COMPETENCE COMMITTEE IMPLEMENTATION AND DECISION MAKING

UNDERSTANDING COMPETENCE COMMITTEE IMPLEMENTATION AND DECISION-MAKING PRACTICES IN THE ERA OF COMPETENCY-BASED MEDICAL EDUCATION

By ANITA ACAI, M.Sc., B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2021)
Hamilton, Ontario (Psychology)

TITLE: Understanding Competence Committee Implementation and Decision-Making Practices in the Era of Competency-Based Medical Education

AUTHOR: Anita Acai, M.Sc., B.Sc. (McMaster University)

SUPERVISOR: Dr. Ranil R. Sonnadara

Number of Pages: xx, 212

Lay Abstract

Competence committees are groups of experienced health professionals and educators whose job is to determine whether physician learners (i.e., residents) are ready to progress to the next stage of training and responsibility. These committees are relatively new, and as a result, we do not know very much about how they make decisions. Given the importance of competence committees in ensuring that physicians are able to provide high-quality and safe patient care, the purpose of this thesis was to examine competence committee implementation and decision-making practices at a Canadian academic centre. This took place in two parts. First, we studied competence committees over a three-year period using surveys, interviews, and observations. This helped us understand some of their benefits and challenges. Next, we conducted a series of experiments to understand how competence committee members make decisions both individually and as part of a group. These experiments also helped us understand how competence committees make sense of different types of data, such as prior knowledge about a resident or their assessors. Finally, we examined how various aspects of members' social identities, such as their position on the committee, their gender, and their race/ethnicity, influence their contributions to the committee. Collectively, the findings of this thesis help to advance the scientific literature in the areas of medical education and group decision making. They can also be used to optimize competence committee operations, which can in turn positively impact patients, healthcare, and society.

Abstract

Competence committees are groups of educators that monitor the progress of medical trainees and decide when they should be promoted to the next stage of training. They represent an important part of modern-day competency-based medical education programs, yet relatively little is known about their implementation and decision-making practices. This thesis seeks to fill a critical gap in the literature by generating empirical evidence with respect to competence committee implementation and decision-making practices across multiple programs. The first data chapter uses a multi-method approach to examine competence committee implementation practices at a Canadian institution over a three-year period. The second and third chapters examine how individuals and groups make promotion decisions, respectively. These chapters also consider the role of non-traditional data sources, such as anecdotal evidence, in competence committees' decision-making processes. The final data chapter considers the role of social influences and power and examines how factors such as members' position on the committee, gender, and race/ethnicity influence their contributions to the committee. This thesis provides insight into some of the challenges that exist with respect to competence committee implementation and offers potential solutions based on best practices across multiple programs. It also highlights factors that can influence competence committee decision making and discusses ways that their decision-making processes can be optimized. Broader implications of this thesis, including the role of groups in solving complex problems and the importance of diversity (both in terms of demographics and functional specialization) in ensuring good decision-making outcomes, are also discussed.

Acknowledgements

The impact of a doctoral degree extends far beyond the words on a page or in a book. As I reflect on my educational journey over the past four and a half years, I am reminded of many memories—many wonderful but a few deeply challenging—that have shaped me into the person I am today. One thing that has remained constant, however, is the village of people that have surrounded and supported me every step of the way. To each of you: thank you. This journey would not have been possible without you.

First and foremost, I would like to thank my supervisor, Dr. Ranil Sonnadara, for giving me a truly outstanding graduate experience. Your support in allowing me to pursue my own research interests, as well as the many professional development opportunities that you have given me, have been immensely helpful in helping me get the most out of my time as a Ph.D. student. Thank you, as well, for your continued support and guidance as I navigated my doctoral journey along its many twists and turns. I know that I will carry all that I have learned from you with me as I embark on the next stage of my career.

I would also like to extend my gratitude to my committee members. Your commitment and service are sincerely appreciated, as are the insights that you have provided along the way. To the Department of Surgery, thank you for supporting this work and my graduate education. Likewise, thank you to the many individuals, both within and outside of the Department, who took the time to participate in my research.

To members of the Sonnadara Lab, both past and present, thank you for creating a supportive place to call home throughout my graduate education. Although I cannot possibly list all of you here, I am grateful to have crossed paths with each of you. A

special thank you to Natalie, Kiret, Tina, Portia, Sydney, Victoria, Lauren, Cindy, Nathan, Ruby, Stella, Maverick, Jeni, and Jacqueline for your continued friendship and support over the years. Ginet, thanks for being a great office mate throughout my doctoral studies—your warmth, kindness, and snacks were always appreciated. Thank you, as well, to friends and colleagues from the Department of Psychology, Neuroscience & Behaviour including Brendan, Hanae, Kaian, Lisa, and Mike, as well as many others who always made me feel at home despite our differing research interests.

To my many collaborators and mentors, both locally and abroad, thank you for providing me with an opportunity to explore my interests both within and outside of health professions education. Among many others, I am especially grateful for the opportunity to have worked with and gotten to know Nancy, Beth, Arshad, Lucy, Rachel, Sam, Teresa, Karen, Sheila, Melissa, and Jillian. Working with each of you has expanded my horizons as a graduate student and brought me a great deal of joy and fulfillment.

Last but not least, thank you, Dad and Mom, for giving me the opportunity to pursue my dreams and taking a deep interest in all that I do. You have been my life's greatest teachers, and I am forever indebted to you for the sacrifices that you have made to build and sustain a strong foundation for our family. I will always love and take care of you both. To all of my friends—especially Jen, Walker, Joey, Eddie, Mary, Lindsay, Gabby, and Paige—thank you for all of the great memories over the years and always being there for me. I look forward to all that lies ahead in our lives.

Table of Contents

CHAPTER 6: GENERAL DISCUSSION

APPENDICES

List of Figures and Tables

CHAPTER 2: A THREE-YEAR, MULTI-PHASE STUDY OF COMPETENCE COMMITTEE IMPLEMENTATION PRACTICES

CHAPTER 3: UNDERSTANDING HOW INDIVIDUAL RATERS MAKE PROMOTION DECISIONS

CHAPTER 5: THE ROLE OF SOCIAL HIERARCHIES AND POWER IN COMPETENCE COMMITTEE DECISION MAKING

List of Abbreviations

CBD	Competence by Design
CBME	Competency-Based Medical Education
CCs	Competence Committees (Canada)
CCCs	Clinical Competency Committees (USA)
EPAs	Entrustable Professional Activities
RCPSC	Royal College of Physicians and Surgeons of Canada
RPC	Residency Program Committee
SDS	Social Decision Scheme

Declaration of Academic Achievement

All of the work in this thesis was conceptualized and carried out by Anita Acai, with input from her supervisor and members of her supervisory committee.

Any exceptions to this are outlined below:

Chapter 2: Nathan Cupido and Aliana Weavers assisted with the transcription and analysis of interviews in Phase. Karen Saperson, Moyez Ladhani, Sharon Cameron, and Elizabeth DeGrow from the McMaster Postgraduate Medical Education office assisted with the development, programming, and deployment of the survey in Phase 2. Jillian Lopes and Jihyun Sung assisted with the transcription of interviews in Phase 3.

This study has been submitted, but not yet accepted, for publication in a peer-reviewed journal. A citation is provided below:

Acai A., Cupido, N., Weavers, A., Saperson, K., Ladhani, M., Cameron, S., & Sonnadara, R. R. (2020). *A three-year, multi-phase study of competence committee implementation practices.* Manuscript in submission.

Chapter 3: Ruby Mann assisted with data collection in Experiment 1. Tianna Murray served as a second coder in Experiment 3.

Chapter 4: Ruby Mann assisted with data collection in Experiment 1. Portia Kalun assisted with collecting observational data from one of the four competence committees included in this study. Elizabeth Clow assisted with the transcription of competence committee meetings.

Chapter 5: Jillian Lopes and Jihyun Sung assisted with the transcription of interviews. Elizabeth Clow assisted with the transcription of competence committee meetings.

Chapter 1

General Introduction

1.1 Competency-Based Medical Education

The global shift from time-based to competency-based models of medical education has created an impetus to reconsider assessment practices in the field (Holmboe, Sherbino, Long, Swing, & Frank, 2010). Competency-based medical education (CBME) requires that trainees do more than just *know how*; they must also *show how* through continuous and frequent assessment based on direct observation in the workplace (Holmboe et al., 2010; Miller, 1990). Under CBME, assessment is not only a method of tracking progress, but also a tool aimed at fostering learning and skill development (i.e., assessment for learning; Holmboe et al., 2010; Lockyer et al., 2017).

Many residency programs have already begun making the shift towards more frequent, workplace-based assessments. This includes the development and implementation of Entrustable Professional Activities (EPAs), which represent the core skills of a given discipline and help guide entrustment decisions (ten Cate, 2013). Another important change brought forth by CBME is the reconceptualization of how competence judgments are made (Holmboe et al., 2010). Rather than having individuals such as program directors make competence judgments based on a small number of end-of-rotation evaluations, these decisions will now be based on a variety of data sources and involve the collective judgment of groups known as competence committees (CCs; Holmboe et al., 2010; Royal College of Physicians and Surgeons of Canada, n.d.-b).

1.2 Competence Committees

CCs (also known as clinical competency committees [CCCs] in the United States) are groups of educators that monitor the progress of medical trainees and decide when they should be promoted to the next stage of training (Hauer et al., 2015; Royal College of Physicians and Surgeons of Canada, n.d.-b). These decisions are based on patterns of performance gleaned from multiple data sources, including EPA observations; documented feedback from clinical practice; exam scores; attendance at academic events; and the completion of specific curricular milestones, such as certifications and research projects (Royal College of Physicians and Surgeons of Canada, n.d.-b). More recently, some researchers have also observed the use of undocumented (i.e., anecdotal) evidence in CC decision-making processes (Ekpenyong et al., 2017; Pack, Lingard, Watling, Chahine, & Cristancho, 2019; Schumacher et al., 2018a).

As part of their mandate, CCs engage in regular, formative reviews of resident progress. This enables the early identification of trainees who are not meeting learning goals and may require support, as well as trainees who are performing well and would benefit from additional learning opportunities (Hauer et al., 2015; Royal College of Physicians and Surgeons of Canada, n.d.-b). At specific points in a resident's training, CCs also make summative judgments about whether they are ready to progress to the next level of training and, eventually, independent practice (Royal College of Physicians and Surgeons of Canada, n.d.-b). In Canada, these judgments result in a recommendation to the residency program committee, which must ratify the decision before it is considered final (Royal College of Physicians and Surgeons of Canada, 2018).

To date, CCs have been implemented in Canada (Royal College of Physicians and Surgeons of Canada, n.d.-b), parts of Europe (Duitsman et al., 2019; General Medical Council, n.d.), and the United States (Andolsek, Padmore, Hauer, Edgar, & Holmboe, 2020). In Canada, CC implementation is mandated as part of the Royal College of Physicians and Surgeons of Canada's (n.d.-b) Competence by Design initiative. Licensure bodies have also mandated CCs or other, similar committees in the Netherlands (Duitsman et al., 2019), United Kingdom (General Medical Council, n.d.), and United States (Andolsek et al., 2020). Some undergraduate programs have also adopted similar approaches (Frank, O'Sullivan, Mills, Muller-Juge, & Hauer, 2019; Monrad et al., 2019).

Regardless of the national context and level of training, the primary role of CCs is similar: to engage in regular reviews of trainees' progress and, at specific time points, determine their readiness to progress to the next stage of training and responsibility. Given this mandate, the decisions made by CCs have the potential to considerably impact a variety of stakeholders. These include trainees, whose educational experiences are in part determined by the recommendations made by CCs, as well residency programs, which have a social and legal obligation produce competent medical professionals capable of providing safe and high-quality patient care (Tweed & Wilkinson, 2019). The reputation of residency programs and healthcare institutions may also be affected by the extent to which their promotion and review processes are seen as credible and defensible (Tweed & Wilkinson, 2019). These features, in combination with the fact that CC implementation is now mandatory in a number of countries across the world, make it important to study how these committees function in practice.

1.3 Existing Literature on Competence Committees

Early literature on CCs has focused primarily on the development of guidelines and theoretical frameworks based on literature from other domains. A review of the group decision-making literature by Hauer and colleagues (2016) revealed five concepts thought to be central to the work of CCs: member characteristics, group understanding of its work, group leader role, information-sharing procedures, and the effects of time pressures. Specific recommendations were provided in each of these areas and have since been expanded upon in other articles to suit a variety of needs, including creating guidelines for virtual promotion meetings (Acai, Sonnadara, & O'Neill, 2018) and providing practical tips for implementing CCs (Edgar & Holmboe, 2019; Kinnear, Warm, & Hauer, 2018). Another review by Chahine, Cristancho, Padgett, and Lingard (2017) examined how various theoretical 'orientations' (i.e., schema, constructivist, and social influence) and moderators (i.e., guidelines, stressors, authority, and leadership) can be used to better understand how CCs and other small groups make decisions.

Still other reviews have explored the role of cognitive biases (Dickey, Thomas, Feroze, Nakshabandi, & Cannon, 2017), conversations (Hemmer & Kelly, 2017), and trust (Sapp, Torre, Larsen, Holmboe, & Durning, 2019) as they might apply to CC decision-making processes. While these articles offer important insights, a considerable shortcoming is that they do not contain any empirical data about how CCs function in practice. This limits their applicability to medical education, since it is difficult to discern whether or not findings from other domains would necessarily apply in the same manner.

A second area of focus within the literature has been to document CC implementation practices across various residency training programs. A variety of case studies exist, including in anesthesiology (Black et al., 2019), internal medicine (Donato, Alweis, & Wenderoth, 2016), pediatrics (Duitsman et al., 2019), and surgery (French, Dannefer, & Colbert, 2014). Rather than testing specific hypotheses, these articles aim to describe how CCs have been implemented within a specific program. With the exception of Hauer and colleagues (2015), who examined CCC implementation practices across California, and Doty and colleagues (2015), who surveyed emergency medicine programs across the United States on the same topic, few articles have considered how CCs may vary across programs.[1] Both Hauer and colleagues' (2015) and Doty and colleagues' (2015) studies revealed considerable variability in CCC structure and function across contexts. This can be problematic because a lack of standardization could be interpreted by learners or the public as there being variable standards across training programs. Thus, there is a need to better understand how CC implementation practices compare across programs.

A third area of focus, which has gained traction in the literature as CC implementation has become more widespread, is to better understand their decision-making practices. A better understanding of how CCs make decisions is important for preventing or addressing potential decision-making pitfalls, informing training, and promoting the credibility and defensibility of training programs. Furthermore, CCs require a substantial

[1] The study by Black and colleagues (2019) included case studies from multiple institutions, but this was not the focus of the article as no comparisons were made between sites.

time commitment from clinical faculty; thus, optimizing their decision-making practices is important to ensure that the time away from providing clinical care is justified.

One of the first studies to examine CC decision-making practices was a study of an internal medicine CCC by Ekpenyong and colleagues (2017), which investigated how members weight different data sources when making promotion decisions. Survey data revealed that resident rotation ratings were given the highest weight (37%), followed by faculty rotation comments (27%), and personal experience with residents (14%). Qualitative findings revealed challenges related to the design of the assessment system (i.e., data quantity and quality), synthesis of assessment data (i.e., CCC decision-making processes), and impact on stakeholders (i.e., how judgments made by CCCs are used)..

Other research related to CC decision-making processes has been conducted in pediatrics by Schumacher and colleagues. A correlational study uncovered a number of factors that may influence how CCC members make summative decisions, with examples being the number of residents reviewed and the number of CCC meetings attended (Schumacher et al., 2018a). Other studies by this group have qualitatively examined the factors that CCC members consider when recommending residents to a supervisory role (Schumacher et al., 2019a) and how their reasoning might differ from that of program directors when making these decisions (Schumacher et al., 2019b). A fourth study considered the processes by which CCCs identify residents with performance concerns (Schumacher et al., 2018b). Residents with performance concerns became apparent in different ways, including through written comments from rotation assessments, concerning performance extremes, isolated data points that accumulate over time, and the

resident's developmental trajectory (e.g., failing to progress or a sudden drop in performance). In interpreting assessment data, it was important for CCC members to assess the quality of the data being reviewed, as some data sources were considered to be more trustworthy than others (Schumacher et al., 2018b).

More recently, Pack, Lingard, Watling, Chahine, and Cristancho (2019) collected observational and interview data from CCs across multiple programs to examine how they made sense of different types of assessment data. Like Schumacher and colleagues (2018b), their findings revealed that the type of decision making that committees engaged in depended on the quality of the assessment data available to them. When CCs were faced with data that were difficult to interpret, they engaged in effortful decision characterized by lengthy discussion and reliance on contextual information. However, when data were congruent and logical, decision making became almost effortless. A second study later revealed that in addition to interpreting assessment data, CCs also engaged in learner-level and program-level problem solving to enhance the quality of their own decision-making processes and provide residents with opportunities for development (Pack, Lingard, Watling, & Cristancho, 2020).

Odorizzi and colleagues (2020) have also conducted research on CC decision-making processes. Here, the authors were interested in understanding how professionalism concerns impact CC decision-making processes. Unlike Schumacher and colleagues (2018b), the authors used an experimental paradigm in which they constructed simulated resident portfolios containing hypothetical formative and summative assessments. Half of the portfolios contained a professionalism variable, which offered additional information

about a resident's professionalism, while the other half did not. The presence of professionalism information led to significantly less consistency among CC members' judgments than when it was not present, leading the authors to conclude that not only did information about professionalism appear to be important in CC decision-making processes, but the variability commonly observed among individual assessors (see for example Gingerich, Kogan, Yeates, Govaerts, & Holmboe, 2014) is also apparent at the level of CC members (i.e., 'meta-raters').

1.4 Gaps in the Literature

To date, the majority of existing scholarship on CCs has focused on topics related to implementation—for example membership, meeting frequency, understanding of their mandate, and the role of academic advisors in supporting their work. With the exception of a handful of studies, the focus of the existing literature has been on CC implementation practices in individual programs. In addition, much of the literature on CCs has originated from the United States, where the implementation of CCCs was mandated by the Accreditation Council for Graduate Medical Education in 2013 (for comparison, mandated implementation in Canada only began in 2017; Andolsek et al., 2020; Royal College of Physicians and Surgeons of Canada, n.d.-a). Thus, there is an opportunity for future studies to consider how CCs have been implemented in different contexts—for example, across programs, institutions, and countries. Furthermore, since existing case studies typically offer a window into implementation practices at only a single point in time, exploring how CCs' needs evolve over time would allow researchers to better understand the challenges that CCs may face at different points in their operations.

While in recent years some studies have begun to explore how CCs make decisions, much of this work has been theoretical or conducted within a single specialty. Additional empirical research on CC decision-making practices is particularly critical given the prevalence of CCs and their potential to impact a wide range of stakeholders. Several studies have shown that, in addition to more conventional data sources such as performance data from EPAs, CCs also use anecdotal evidence (i.e., undocumented information about a resident or their assessors) when making promotion decisions (Ekpenyong et al., 2017; Hauer et al., 2015; Pack et al., 2019; Schumacher et al., 2018b). However, in comparison to EPA data, little is known about how this type of information is used by CCs, and why. This question is particularly interesting in light of the fact that the literature often positions these committees as relying on more 'objective' data sources than previous promotion and review processes (Pack et al., 2019), which is in contradiction to the idea that CCs might use—and even benefit from—subjective data such as anecdotal evidence when making promotion decisions.

Another question relates to the role of social hierarchies in CC decision making. A theoretical review of the group decision-making literature by Hauer and colleagues (2016) revealed member characteristics to be an important factor in CC decision-making processes. For example, members in a leadership role—such as the chair or program director—may be particularly influential in the committee's decision-making processes (Chahine et al., 2017; Hauer et al., 2016). The social psychology literature also suggests that there may be differences in members' contributions based on their position, gender, or race/ethnicity (Elsass & Graves, 1997). Understanding the role of social influences is

important for understanding how CCs operate but may also have broader applications. For example, the emerging concept of collective competence suggests that healthcare outcomes are often dependent on the ability of multiple healthcare providers to work together in a complex environment (Sebok-Syer et al., 2018). Thus, understanding how social hierarchies impact decision making in CCs may also be relevant to other group processes in medicine, such as those related to assessment or entrustment (Sebok-Syer et al., 2018). Moreover, considering who is represented on decision-making bodies and how this informs decision making is also important for addressing issues related to equity in medicine, which is critical for ensuring that healthcare systems can adequately serve the communities they represent, as well as for creating positive educational and workplace environments (Canadian Medical Association, 2019; Price et al., 2009; Yancy, 2020).

1.5 The Present Thesis

The present thesis seeks to fill a critical gap in the literature by generating empirical evidence with respect to CC implementation and decision-making practices across multiple programs. Using a multi-method approach that draws upon surveys, interviews, observations, and experiments, this thesis begins with a longitudinal study of CC implementation of CCs at a Canadian institution followed by a theoretically grounded examination of their decision-making practices. A particular area of focus is how CCs use anecdotal evidence and how it may inform the interpretation of other data. Collectively, the aim of this thesis is to provide a better understanding of the extent to which CCs are meeting the mandate put forth by CBME, and the ways in which they can continue to be

optimized to ensure robust educational outcomes for residency training programs. Broader implications of this thesis (i.e., beyond CCs) will also be explored.

The national transition to CBME in Canada, which officially began in 2017 and will continue on a rolling basis until the end of 2025, presents a unique opportunity to study CC implementation as it unfolds across the country (Royal College of Physicians and Surgeons of Canada, n.d.-a). Chapter 2 of this thesis presents a study of CC implementation at a Canadian institution that documents implementation practices longitudinally. This study was conducted in three phases over three years and consisted of a multi-method approach comprising interviews, observations, and surveys. This study explores the perceived benefits and challenges of CC implementation, and highlights a number of areas for further investigation, including how CCs make decisions.

The remainder of this thesis is focused on developing a better understanding of how CCs make promotion decisions. Chapter 3 considers the role of individual raters in this process through three related experiments. Recognizing that most decision-making tasks are inextricable from the context in which they occur, the first experiment examines how novice raters (i.e., undergraduate students) make promotion decisions using simulated resident portfolios containing different types of anecdotal evidence that contain undocumented but potentially important contextual information about a resident's performance and/or circumstances. The second experiment is a replication of the first using a sample of experienced raters (i.e., current CC members with approximately two years of experience). The third experiment replicates the first and involves an analysis of novice raters' written reasoning for their promotion decisions.

Chapter 4 considers the role of groups in making promotion decisions, and in particular, how group decisions compare to the individual decisions presented in Chapter 3. Once again, three related experiments/analyses are presented. The first uses the same stimuli as in Chapter 3 to examine how groups of novice raters make promotion decisions involving anecdotal evidence. The second is a replication using groups of experienced raters (i.e., existing CCs). The third involves an in-depth analysis of interviews with CC members and meeting transcripts to examine the role that anecdotal evidence plays in CC decision-making processes in naturalistic environments.

Chapter 5 considers the role of social hierarchies and power in CC decision-making processes. First, an analysis of CC meeting transcripts is conducted to examine how the number of speaking turns and words spoken vary based on members' position, gender, and race/ethnicity. Next is an analysis of promotion decisions made by CCs, including the characteristics of members contributing influential information. In the final analysis, members' perceptions of social hierarchies and power are analyzed through interviews conducted with junior and senior CC members.

The final chapter, Chapter 6, serves as a general discussion, integrating the findings presented in the previous chapters, exploring their implications, and providing recommendations for future research.

1.6 References

Acai, A., Sonnadara, R. R., & O'Neill, T. A. (2018). Getting with the times: A narrative review of the literature on group decision making in virtual environments and

implications for promotions committees. *Perspectives on Medical Education, 7*(3), 147-155. https://doi.org/10.1007/s40037-018-0434-9

Andolsek, K., Padmore, J., Hauer, K. E., Edgar, L., & Holmboe, E. (2020). Clinical competency committees: A guidebook for programs (3rd edition). Retrieved from https://www.acgme.org/Portals/0/ACGMEClinicalCompetencyCommitteeGuidebook.pdf

Black, S., Capdeville, M., Augoustides, J. G. T., Nelson, E. W., Patel, P. A., Feinman, J. W., . . . Yanofsky, S. D. (2019). The clinical competency committee in adult cardiothoracic anesthesiology—perspectives from program directors around the United States. *Journal of Cardiothoracic and Vascular Anesthesia, 33*(7), 1819-1827. https://doi.org/10.1053/j.jvca.2019.01.001

Canadian Medical Association. (2019). CMA policy: Equity and diversity in medicine. Retrieved from https://policybase.cma.ca/documents/Policypdf/PD20-02.pdf

Chahine, S., Cristancho, S., Padgett, J., & Lingard, L. (2017). How do small groups make decisions? A theoretical framework to inform the implementation and study of clinical competency committees. *Perspectives on Medical Education, 6*(3), 192-198. https://doi.org/10.1007/s40037-017-0357-x

Dickey, C. C., Thomas, C., Feroze, U., Nakshabandi, F., & Cannon, B. (2017). Cognitive demands and bias: Challenges facing clinical competency committees. *Journal of Graduate Medical Education, 9*(2), 162-164. https://doi.org/10.4300/JGME-D-16-00411.1

Donato, A. A., Alweis, R., & Wenderoth, S. (2016). Design of a clinical competency committee to maximize formative feedback. *Journal of Community Hospital Internal Medicine Perspectives, 6*(6). https://doi.org/10.3402/jchimp.v6.33533

Doty, C. I., Roppolo, L. P., Asher, S., Seamon, J. P., Bhat, R., Taft, S., . . . Willis, J. (2015). How do emergency medicine residency programs structure their clinical competency committees? A survey. *Academic Emergency Medicine, 22*(11), 1351-1354. https://doi.org/10.1111/acem.12804

Duitsman, M. E., Fluit, C. R. M. G., van Alfen-van der Velden, J. A. E. M., de Visser, M., ten Kate-Booij, M., Dolmans, D. H. J. M., . . . de Graaf, J. (2019). Design and evaluation of a clinical competency committee. *Perspectives on Medical Education, 8*(1), 1-8. https://doi.org/https://doi.org/10.1007/s40037-018-0490-1

Edgar, L., & Holmboe, E. (2019). International clinical competency committees: Maximizing value for faculty, residents, and the program. *Journal of Graduate Medical Education, 11*, 191-192. https://doi.org/10.4300/JGME-D-19-00413

Ekpenyong, A., Baker, E., Harris, I., Tekian, A., Abrams, R., Reddy, S., & Park, Y. S. (2017). How do clinical competency committees use different sources of data to assess residents' performance on the internal medicine milestones? A mixed methods pilot study. *Medical Teacher, 39*(10), 1074-1083. https://doi.org/10.1080/0142159X.2017.1353070

Elsass, P. M., & Graves, L. M. (1997). Demographic diversity in decision-making groups: The experiences of women and people of color. *The Academy of Management Review, 22*(4), 946-973. https://doi.org/10.2307/259250

Frank, A. K., O'Sullivan, P., Mills, L. M., Muller-Juge, V., & Hauer, K. E. (2019). Clerkship grading committees: The impact of group decision-making for clerkship grading. *Journal of General Internal Medicine, 34*(5), 669-676. https://doi.org/https://doi.org/10.1007/s11606-019-04879-x

French, J. C., Dannefer, E. F., & Colbert, C. Y. (2014). A systematic approach toward building a fully operational clinical competency committee. *Journal of Surgical Education, 71*(6), e22-27. https://doi.org/10.1016/j.jsurg.2014.04.005

General Medical Council. (n.d.). Annual review of competency progression (ARCP). Retrieved from https://www.gmc-uk.org/education/reports-and-reviews/progression-reports/annual-review-of-competency-progression

Gingerich, A., Kogan, J., Yeates, P., Govaerts, M., & Holmboe, E. (2014). Seeing the 'black box' differently: Assessor cognition from three research perspectives. *Medical Education, 48*(11), 1055-1068. https://doi.org/10.1111/medu.12546

Hauer, K. E., Chesluk, B., Iobst, W., Baron, R. B., Boscardin, C. K., ten Cate, O., & O'Sullivan, P. S. (2015). Reviewing residents' competence: A qualitative study of the role of clinical competency committees in performance assessment. *Academic Medicine, 90*(8), 1084-1092. https://doi.org/10.1097/ACM.0000000000000736

Hauer, K. E., ten Cate, O., Boscardin, C. K., Iobst, W., Holmboe, E. S., Chesluk, B., . . . O'Sullivan, P. S. (2016). Ensuring resident competence: A narrative review of the literature on group decision making to inform the work of clinical competency committees. *Journal of Graduate Medical Education, 8*(2), 156-164. https://doi.org/10.4300/JGME-D-15-00144.1

Hemmer, P. A., & Kelly, W. F. (2017). We need to talk: Clinical competency committees in the key of c(onversation). *Perspectives on Medical Education, 6*(3), 141-143. https://doi.org/10.1007/s40037-017-0360-2

Holmboe, E. S., Sherbino, J., Long, D. M., Swing, S. R., & Frank, J. R. (2010). The role of assessment in competency-based medical education. *Medical Teacher, 32*(8), 676-682. https://doi.org/10.3109/0142159X.2010.500704

Kinnear, B., Warm, E. J., & Hauer, K. E. (2018). Twelve tips to maximize the value of a clinical competency committee in postgraduate medical education. *Medical Teacher, 40*(11), 1110-1115. https://doi.org/10.1080/0142159X.2018.1474191

Lockyer, J., Carraccio, C., Chan, M.-K., Hart, D., Smee, S., Touchie, C., . . . ICBME Collaborators. (2017). Core principles of assessment in competency-based medical education. *Medical Teacher, 39*(6), 609-616. https://doi.org/10.1080/0142159X.2017.1315082

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*, S63-S67. https://doi.org/10.1097/00001888-199009000-00045

Monrad, S. U., Mangrulkar, R. S., Woolliscroft, J. O., Daniel, M. M., Hartley, S. E., Gay, T. L., . . . Santen, S. A. (2019). Competency committees in undergraduate medical education: Approaching tensions using a polarity management framework. *Academic Medicine, 94*(12), 1865-1872. https://doi.org/https://doi.org/10.1097/ACM.0000000000002816

Odorizzi, S., Cheung, W. J., Sherbino, J., Lee, A. C., Thurgur, L., & Frank, J. R. (2020). A signal through the noise: Do professionalism concerns impact the decision-making of competence committees? *Academic Medicine, 95*(6), 896-901. https://doi.org/10.1097/ACM.0000000000003005

Pack, R., Lingard, L., Watling, C. J., Chahine, S., & Cristancho, S. M. (2019). Some assembly required: Tracing the interpretive work of clinical competency committees. *Medical Education, 53*(7), 723-734. https://doi.org/10.1111/medu.13884

Pack, R., Lingard, L., Watling, C. J., & Cristancho, S. (2020). Beyond summative decision-making: Illuminating the broader roles of competence committees. *Medical Education, 54*(6), 517-527. https://doi.org/10.1111/MEDU.14072

Price, E. G., Powe, N. R., Kern, D. E., Golden, S. H., Wand, G. S., & Cooper, L. A. (2009). Improving the diversity climate in academic medicine: Faculty perceptions as a catalyst for institutional change. *Academic Medicine, 84*(1), 95-105. https://doi.org/10.1097/ACM.0b013e3181900f29

Royal College of Physicians and Surgeons of Canada. (2018). Competence committee guideline: Process and procedures in decision making. Retrieved from http://www.royalcollege.ca/rcsite/documents/cbd/competence-committees-process-procedures-e

Royal College of Physicians and Surgeons of Canada. (n.d.-a). Competence by Design launch schedule. Retrieved from http://www.royalcollege.ca/rcsite/cbd/cbd-implementation-e

Royal College of Physicians and Surgeons of Canada. (n.d.-b). Competence committees. Retrieved from http://www.royalcollege.ca/rcsite/cbd/assessment/competence-committees-e

Sapp, J. E., Torre, D. M., Larsen, K. L., Holmboe, E. S., & Durning, S. J. (2019). Trust in group decisions: A scoping review. *BMC Medical Education, 19*(1), 309. https://doi.org/10.1186/s12909-019-1726-4

Schumacher, D. J., King, B., Barnes, M. M., Elliott, S. P., Gibbs, K., McGreevy, J. F., . . . Appd Learn CCC Study Group. (2018a). Influence of clinical competency committee review process on summative resident assessment decisions. *Journal of Graduate Medical Education, 10*(4), 429-437. https://doi.org/10.4300/JGME-D-17-00762.1

Schumacher, D. J., Martini, A., Bartlett, K. W., King, B., Calaman, S., Garfunkel, L. C., . . . Appd Learn CCC Study Group. (2019a). Key factors in clinical competency committee members' decisions regarding residents' readiness to serve as supervisors: A national study. *Academic Medicine, 94*(2), 251-258. https://doi.org/10.1097/ACM.0000000000002469

Schumacher, D. J., Michelson, C., Poynter, S., Barnes, M. M., Li, S.-T. T., Burman, N., . . . Appd Learn CCC Study Group. (2018b). Thresholds and interpretations: How clinical competency committees identify pediatric residents with performance concerns. *Medical Teacher, 40*(1), 70-79. https://doi.org/10.1080/0142159X.2017.1394576

Schumacher, D. J., Poynter, S., Burman, N., Elliott, S. P., Barnes, M., Gellin, C., . . . Appd Learn CCC Study Group. (2019b). Justifications for discrepancies between competency committee and program director recommended resident supervisory roles. *Academic Pediatrics, 19*(5), 561-565. https://doi.org/10.1016/j.acap.2018.12.003

Sebok-Syer, S. S., Chahine, S., Watling, C. J., Goldszmidt, M., Cristancho, S., & Lingard, L. (2018). Considering the interdependence of clinical performance: Implications for assessment and entrustment. *Medical Education, 52*(9), 970-980. https://doi.org/10.1111/medu.13588

ten Cate, O. (2013). Nuts and bolts of Entrustable Professional Activities. *Journal of Graduate Medical Education, 5*(1), 157-158. https://doi.org/10.4300/JGME-D-12-00380.1

Tweed, M., & Wilkinson, T. (2019). Student progress decision-making in programmatic assessment: Can we extrapolate from clinical decision-making and jury decision-making? *BMC Medical Education, 19*. https://doi.org/https://doi.org/10.1186/s12909-019-1583-1

Yancy, C. W. (2020). Academic medicine and Black Lives Matter: Time for deep listening. *Journal of the American Medical Association*. https://doi.org/10.1001/jama.2020.12532

Chapter 2

A Three-Year, Multi-Phase Study of Competence Committee Implementation Practices

2.1 Introduction

In 2015, the Royal College of Physicians and Surgeons of Canada (RCPSC) announced that all residency training programs in Canada would transition to a new competency-based curriculum known as Competence by Design (CBD; Royal College of Physicians and Surgeons of Canada, 2014). As part of this transition, the RCPSC mandated the implementation of competence committees (CCs; also known as clinical competency committees [CCCs] in the United States), groups of educators tasked with reviewing residents' progress and making promotion decisions based on a variety of data sources (Royal College of Physicians and Surgeons of Canada, n.d.-b). Although CCs and other, similar committees have existed in the United States (Andolsek, Padmore, Hauer, Edgar, & Holmboe, 2020; Monrad et al., 2019) and parts of Europe (Duitsman et al., 2019; General Medical Council, n.d.) for some time, there has not been a comprehensive, multi-program study of CC implementation practices since 2015, which predates the implementation of most CCs outside of the United States (Hauer et al., 2015). This represents a critical gap due to the important roles that CCs play in facilitating trainee development and overseeing their progression towards competence, which, in turn, can impact patients, healthcare, and society (Tweed & Wilkinson, 2019).

To date, the majority of the literature on CCs has documented CC design and implementation within individual programs (Donato, Alweis, & Wenderoth, 2016;

Duitsman et al., 2019; Ketteler et al., 2014) or drawn on literature from other domains to offer guidance to programs implementing CCs (Acai, Sonnadara, & O'Neill, 2018; Chahine, Cristancho, Padgett, & Lingard, 2017; Dickey, Thomas, Feroze, Nakshabandi, & Cannon, 2017; Edgar & Holmboe, 2019; French, Dannefer, & Colbert, 2014; Hauer et al., 2016; Hemmer & Kelly, 2017; Kinnear, Warm, & Hauer, 2018). Some studies have also begun to explore aspects of the CC decision-making processes (Odorizzi et al., 2020; Pack, Lingard, Watling, Chahine, & Cristancho, 2019; Schumacher et al., 2018a; Schumacher et al., 2019a; Schumacher et al., 2018b; Schumacher et al., 2019b) and shed light on their role in solving problems at the learner and program levels (Pack, Lingard, Watling, & Cristancho, 2020). Fewer studies have collected empirical data on CC implementation practices, with those that have tending to focus only on a single specialty. For example, a study by Doty et al. (2015) surveyed emergency medicine CCCs across the United States and found considerable variability in how they were implemented. While some programs included nurses (5%) or residents (25%) as members, others did not. There were also differences in how many residents were reviewed at each meeting and for how long, as well as the tasks that CCCs viewed as part of their mandate. More recently, a single-site study of an internal medicine CCC by Ekpenyong et al. (2017) identified challenges related to the design of the assessment system (i.e., data quantity and quality), synthesis of assessment data (i.e., CCC decision-making processes), and impact on stakeholders (i.e., how judgments made by CCCs are used).

Perhaps the most comprehensive assessment of CCC implementation practices to date has been by Hauer et al. (2015), who conducted interviews with 34 residency program

directors at five public institutions in California. Collectively, these individuals represented 22 large and 12 small programs and encompassed 15 procedural and 19 nonprocedural specialties. Like Doty et al. (2015) and Ekpenyong et al. (2017), the authors found considerable variability in CCC implementation practices across programs. Another important finding was that most committees tended to use a problem identification versus a developmental model, meaning that they spent more time reviewing underperforming than average or exceptional residents. The authors also identified limitations in the amount of training provided to members, with only 14 of 21 programs having formally onboarded or provided training to their CCC members.

Since Hauer et al.'s (2015) study was conducted in the United States, it is unclear if the findings are generalizable to other countries in which CBME is being implemented. Furthermore, the study did not investigate if and how CC implementation practices evolve over time, making it difficult to identify and address the challenges that CCs may face at different points in the implementation process. Given the nationally mandated transition to CBD (Royal College of Physicians and Surgeons of Canada, 2014), the Canadian context offers a unique opportunity to document a major transition in medical education as it unfolds across programs and time. The present study explored CC implementation at a Canadian institution over a three-year period, thereby addressing some of the aforementioned gaps in the literature. To the investigators' knowledge, this is the first longitudinal study to have examined CC implementation at any institution.

2.2 Methods

2.2.1 Context and Design

This longitudinal study took place at McMaster University, a mid-sized academic training centre in Hamilton, Ontario, Canada with 57 specialty and sub-specialty programs. It began in 2017 when the majority of residency training programs in Canada were pre-CBD implementation and ended in 2020, when CCs implemented at the onset of CBD had been in operation for at least two years. A multi-method approach involving interviews, surveys, and observations was used to examine CC implementation as it evolved over the study period. The study took place in three phases, detailed below and shown in Figure 1. Since CBD implementation in Canada follows a staggered approach (Royal College of Physicians and Surgeons of Canada, n.d.-a), not all programs implemented CCs at the same time. However, some programs chose to pilot CCs prior to their scheduled transition to the new curriculum.

2.2.2 Phase 1: Pre-Implementation

The first phase of the study took place between May 2017 and February 2018. During this time, most residency training programs in Canada were pre-CBD implementation; thus, most CCs were in the planning stages, and a few pilot committees had been running for a short period of time. Thus, this phase of the study sought to capture key stakeholders' early perceptions and experiences of CC implementation.

A purposive sample of 30 key stakeholders from across the institution were invited to participate in the study. Since a central list of existing CCs and chairs was not yet available, recruitment was conducted using an initial list of faculty and residents

known to be leaders in competency-based medical education, followed by a snowball sampling approach (Mack, Woodsong, MacQueen, Guest, & Namey, 2005). Inclusion criteria for participants were having been involved in resident promotion and review processes within the last year and/or having held leadership roles in postgraduate medicine. These criteria helped ensure that participants had sufficient knowledge and experience to offer meaningful perspectives on CC implementation.

Following consent, participants were sent a 14-item electronic survey to gauge their awareness of CCs; attitudes about CCs; and opinions about membership, existing resources, and workload (Appendix 1). The survey was coded in LimeSurvey by the primary investigator, a PhD candidate in psychology and health professions education. Input on the survey design was sought from clinicians and a senior education scientist, who was also a CC member. Findings were summarized using descriptive statistics.

Participants were also asked if they wished to take part in a brief follow-up interview that would further probe issues related to CC implementation using a qualitative descriptive design, which allowed for a rich description of participants' experiences (Neergaard, Olesen, Anderson, & Sondergaard, 2009). If they agreed, an interview was scheduled using the medium of their choosing (face-to-face or telephone; see Appendix 2 for interview guide). All interviews were audio-recorded, transcribed, and de-identified prior to analysis. Analysis of the interviews took place using Braun and Clarke's (2006) six-step approach to thematic analysis: familiarizing oneself with the data, generating initial codes, searching for themes, reviewing themes, naming themes, and producing the final report. Each interview was coded by at least two investigators and regular meetings

were held to ensure consensus with respect to the coding framework and themes as they evolved. Strategies to ensure the trustworthiness of the data included reflexive journaling, documenting the analytic process, and regular debriefing with experienced clinicians both within and outside of the research team (Nowell, Norris, White, & Moules, 2017).

2.2.3 Phase 2: Early Implementation (One Year Mark)

This phase of the study took place in January 2019. During this time, the majority of CCs included in this study had been in operation for approximately one year and more were being implemented. A 38-item electronic survey coded in SurveyGizmo was developed by the primary investigator in collaboration with a team of clinicians and administrators from McMaster's Postgraduate Medical Education Office (Appendix 3). The purpose of the survey was to develop an in-depth understanding of CC processes, including membership, meeting frequency, orientation practices, and decision-making procedures. Input on the survey design was sought from three CC chairs.

An invitation to participate in the survey was sent to all CC chairs (n = 35) at McMaster University. A follow-up email was sent several weeks later. Quantitative survey data were summarized using descriptive statistics while open-ended responses were reviewed by the primary investigator and grouped into themes.

2.2.4 Phase 3: Late Implementation (Two Year Mark)

This phase took place in June and July 2020, at which time the earliest CCs had been in operation for over two years and many more residency training programs had begun their transition to CBD. During this time, the primary investigator invited 20 CC members to participate in a telephone interview investigating how their perceptions and

experiences of CC implementation had evolved over time. All known resident members of a CC were contacted, along with individuals who had previously expressed an interest in participating in CC-related research. The same data collection and analysis procedures were used as in the qualitative strand of the pre-implementation phase (see Appendix 4 for the Phase 3 interview guide, which was informed by prior findings).

2.2.5 Observations

Between August 2017 and May 2020, the primary investigator observed 16 CC meetings across nine of the specialties that took part in Phase 2 of the study and took detailed notes during each meeting. For two CCs, she attended and recorded three to four meetings per year for a one and two-year period, respectively. For others, she attended only one or two meetings per committee but used these experiences to compare and contrast implementation practices across different programs. Notes from the observations were used to contextualize and triangulate data from the interviews and surveys.

2.2.6 Ethics Approval

This study received approval from the Hamilton Integrated Research Ethics Board (HiREB-4248). Some portions (i.e., the observations and surveys) were deemed exempt from ethics review as they were considered quality improvement.

2.3 Results

2.3.1 Phase 1: Pre-Implementation

Of the 30 individuals contacted, 25 (83%) responded to the invitation to participate. Of these 25 participants, 22 (88%) completed the survey on CC implementation and 24 (96%) participated in an interview. No new themes were

identified after the tenth interview, but the remaining individuals were included to ensure representation across programs and stakeholder groups. Participants included 17 faculty and eight residents affiliated with 13 residency programs. Nine participants were current members of a CC, five were to become members within the next year, one was a past member, and nine were not members but were familiar with the concept through prior involvement in decision making related to resident promotion and review (e.g., by serving on a residency program committee [RPC]). One member reported being unsure of their membership status but did not specify a reason.

Survey responses revealed that the majority of faculty were supportive of CC implementation, with 93% (n = 13) of faculty agreeing or strongly agreeing that they would improve decision-making processes around resident promotion and 71% (n = 10) of faculty agreeing or strongly agreeing that they would improve educational outcomes for residency programs. Support for CCs among residents was lower, with only 63% (n = 5) of residents agreeing or strongly agreeing that they would improve decision-making processes around resident promotion and 50% (n = 4) of residents agreeing or strongly agreeing that they would improve educational outcomes for residency programs.

The survey also probed participants' perspectives on CC membership. Seventy-nine percent (n = 11) of faculty agreed or strongly agreed that faculty external to a residency program should be included as CC members, compared with only 25% (n = 2) of residents. The opposite was true for resident members, who were supported by 75% (n = 6) of residents but only 57% (n = 8) of faculty. Neither faculty (14%; n = 2) nor

residents (13%; *n* = 1) were supportive of having patients and individuals outside of the medical education community serve as CC members.

Although participants were generally positive about CC implementation, their survey responses also revealed some perceived challenges. Ninety-one percent (*n* = 20) of participants believed that CCs would require greater commitment (i.e., time and amount of work) from faculty than current promotion and review processes. Moreover, only 36% (*n* = 8) of participants believed that their program had the appropriate tools and resources to implement CCs, and only 56% (*n* = 5) of current CC members felt that they, as members, had the appropriate tools and resources to carry out their duties effectively.

Qualitative findings corroborated the survey results, highlighting participants' belief that CCs would increase the credibility and defensibility of resident promotion and review processes. However, the interviews also revealed some additional challenges. Large programs were perceived to struggle with workload while small programs with managing preexisting relationships between faculty and residents. Participants also felt that more robust ways of sharing and aggregating data were needed for CCs to function properly, as current methods posed a challenge. A potential source of concern was that some participants, especially resident non-members, described CCs as a 'black box.' While they understood the concept, they were uncertain about the mechanics of how decisions would be made and the impact that this would have on their training experience. Key findings and quotes from the qualitative portion of this phase are provided in Table 1.

2.3.2 Phase 2: Early Implementation (One Year Mark)

Survey responses were received from 15 out of 35 chairs, reflecting a 43% response rate. Of these, 14 programs (93%) reported having a CC, with 11 (79%) of these being fully functional (i.e., meeting and reviewing residents regularly) and the other three meeting for planning and/or training purposes but not yet reviewing residents on a regular basis. Across the programs that reported having CCs, five represented small programs (one to nine residents), five represented mid-sized programs (10 to 39 residents), and four represented large programs (40 or more residents). Using the classification system previously reported by Hauer et al. (2015), six programs represented procedural specialties and eight represented non-procedural specialties.

CC membership ranged between five and nine members. Five programs (36%) reported having at least one resident as a member and nine (64%) reported having at least one external member. External members constituted allied health professionals, faculty members from another program, or PhD-trained researchers (e.g., education scientists). The majority of CCs (71%; $n = 10$) met quarterly and half (50%; $n = 7$) reported providing their members with some form of orientation or training, most often involving a presentation on the CC's role and/or a hands-on activity with simulated resident files. At the time of the survey (i.e., pre-COVID), 79% ($n = 11$) of committees offered teleconferencing for members who could not physically attend; however, use of this option was discouraged unless absolutely required. All CCs reported reviewing residents at least twice per year, which is consistent with RCPSC guidelines.

Themes derived from the open-ended responses included varied perceptions of the academic advisor role, varied perceptions of member workload, challenges with data sharing, challenges with engaging faculty and residents, and varied understanding of CCs' mandate. A majority of CCs (79%; *n* = 11) reported having a CBD academic advisor system in place (sometimes also referred to as academic coaches), yet there was a range in how this role was operationalized. There was no set way of choosing academic advisors, with individuals serving in this role ranging from longitudinal clinical supervisors to faculty who were selected by the program director to CC members themselves, particularly in smaller programs. In some cases, the only engagement was one or two meetings between the resident and the advisor per year whereas, in others, the advisor was regularly invited to attend CC meetings and report on the resident's progress, including the assessment data in their file.

The majority (79%; *n* = 11) of CCs did not have a set time limit for file review and spent as much time as needed on each file. Residents who were not progressing as expected were allotted much more time for review than residents who were performing well. On average, CCs spent approximately five to ten minutes per resident file, but on occasion, could spend up to half an hour discussing residents who were not progressing as expected. The average length of each CC meeting was between one and three hours in length, depending on the size of the program. There were also notable differences in how committees engaged in the review process, in part due to their size. Smaller programs were often able to engage in an in-depth review of each resident file as a committee. However, in larger programs, CCs had to rely on a member assignment system in which

one or two members were responsible for summarizing the assessment data and in each resident file and making a recommendation to the committee. On average, each CC member reviewed between two and six files per meeting.

At the time of the survey, members were using a range of electronic platforms to share data securely. Learning how to fully take advantage of these technologies was a challenge, as was cross-referencing between systems while programs transitioned to a single, institutionally recommended platform. It was also common for CCs to encounter missing or incomplete assessment data during the promotion and review process, as residents did not always trigger assessments and faculty did not always fill them out. Thus, identifying and reporting on areas for faculty and resident development to the RPC became an important part of CCs' roles, although other aspects of CCs' mandate were still being defined. For example, while respondents felt that a clear advantage of CCs was that they helped identify residents in need of support much earlier on in their training than previous processes, programs were still grappling with the extent to which they should be involved in the development of learning and remediation plans. At the time of the survey, just over half (57%; $n = 8$) of CCs believed that providing input into residents' learning and remediation plans was a part of their mandate.

2.3.3 Phase 3: Late Implementation (Two Year Mark)

Of the 20 individuals contacted, 12 (60%) responded to the invitation to participate. No new themes were identified after eight interviews, but the four remaining participants were interviewed to ensure that the perspectives of everyone who responded were captured. Participants represented CCs across seven specialties, and a range of

membership types (clinical faculty: $n = 7$, PhD-trained researchers: $n = 1$, residents: $n = 4$). Seven participants served on only one CC while three served on two or more CCs. Participants had attended an average of six CC meetings at the time of the interview. Key findings and quotes from this phase are provided in Table 2.

Observations and follow-up interviews from CCs in the late phases of implementation revealed that many committees had begun to streamline their operations. Members were increasingly comfortable with the promotion and review process, which led to greater efficiency during meetings. Many programs had also introduced a file assignment system, where each CC member was tasked with reviewing a subset of resident files. Some members felt that a clear advantage of CCs was that they reduced the workload placed on individual committee members, as previous processes required a single individual, usually the program director, to review every resident file on their own. Other members, however, suggested that the workload for CC members was still high. On average, committee members reported spending approximately half an hour per resident on file reviews prior to attending a CC meeting, which could amount to between one and three hours of preparation time depending on the number of files reviewed.

All CC members interviewed as part of this study believed their committees to have a collegial atmosphere in which they felt comfortable sharing their opinions freely. In most cases, this was because committee members had been deliberately selected to ensure that this would be the case (e.g., by selecting an experienced chair, who then helped to populate the committee with members believed to offer useful perspectives and work well with one another). While some programs continued to feel strongly about

residents not being CC members, those that did include trainees reported that they brought an important perspective and were able to learn from the experience. Importantly, residents who had served on CCs considered their experience to be an excellent learning opportunity that gave them confidence in their program's approach to resident promotion and review. Some CCs also found it valuable to include PhD-trained researchers or faculty from other specialties (e.g., those representing off-service rotations) as members, as these individuals could help ensure that the committee was applying evaluation criteria consistently and remained accountable for its actions.

2.4 Discussion

To the investigators' knowledge, this is the first longitudinal study to have examined CC implementation at any institution. Consistent with literature from the United States (Doty et al., 2015; Ekpenyong et al., 2017; Hauer et al., 2015), this study revealed considerable variability in how CCs operate across different programs at a single institution. This is not unexpected, as national guidelines allow for some flexibility in adapting CCs to program-level needs, which is important for successful implementation (Royal College of Physicians and Surgeons of Canada, n.d.-c). However, this flexibility may also feel daunting to those new to CC implementation. The remaining sections of this discussion aim to distill the findings of this study into practical considerations and recommendations for CC implementation.

2.4.1 Optimizing Membership

When deciding on CC membership, external members can provide an outside perspective and help keep the committee accountable for its actions (French et al., 2014;

Kinnear et al., 2018). However, participants in the present study felt strongly that patients and individuals outside of the medical education community should not be included as CC members, as they were perceived to lack the context needed to meaningfully comment on learner performance. Since CCs are ultimately accountable to patients and society, it is possible that some programs may consider involving patient representatives once their committees become more established (Loo, Lee, & Acosta, 2017). This decision should involve thoughtful consideration of both the pros and the cons, as there is a risk of negative consequences if execution is poor (e.g., a loss of patient trust in the healthcare system or poor morale among residents). For programs wishing to include a patient representative, it may be prudent to select someone with prior experience in a healthcare setting, such as service on a hospital committee. For programs that do not feel comfortable including patients as CC members, a compromise may be to include PhD-trained researchers (e.g., education scientists) or faculty from other specialties (Loo et al., 2017). Doing so can provide opportunities to share best practices and build capacity across programs while also keeping CCs accountable for their decisions.

Including residents on CCs may also introduce an important perspective and serve as a valuable professional development opportunity given that medicine is a self-regulated profession (Creuss & Creuss, 2000); however, the appropriateness of doing so ultimately depends on program culture. Programs that are unsure about whether to include residents as members may find it helpful to involve them in deciding about this matter, such as through discussions at the residency program committee level. Programs that do include resident members should ensure that they have clear guidelines for

managing conflicts of interest and confidentiality, while those that do not may wish to consider alternative ways of involving residents, such as through the establishment of a CBD working group to help inform curriculum implementation within the program.

2.4.2 Maintaining Capacity Among Members

This study also identified challenges related to maintaining capacity among CC members, which is consistent with other research on assessment in CBD that reports a longstanding tension between resident education and service provision (Li, Acai, Sherbino, & Chan, in press; McQueen et al., 2016). Strategies for rewarding and managing faculty workload involve assigning merit points for individuals who serve on CCs, assigning a set number of reviewers per file, and having members come to meetings having pre-reviewed their assigned files to increase efficiency (Kinnear et al., 2018). Another strategy may be to periodically rotate CC members where possible (French et al., 2014). Both in the present study and elsewhere (e.g., Donato et al., 2016), some CCs chose to involve academic advisors in file reviews to reduce the burden on individual committee members. Although potentially inevitable in smaller programs, having CC members double as academic advisors is an approach that should be adopted with caution as it can be challenging to ensure that advisors are appropriately trained to interpret assessment data and able to attend CC meetings when required. There may also be an inherent conflict of interest in asking faculty to serve as both advisors and file reviewers, as residents may feel uncomfortable disclosing certain issues to someone they know will be evaluating them (Andolsek et al., 2020). In general, the findings show that the role of

the academic advisor (or coach) is interpreted differently across programs, suggesting a need to clarify how this role is expected to interface with CCs.

2.4.3 Engaging Faculty and Residents in CBD

Residents who were not CC members often had the impression that CCs were a 'black box.' This is consistent with a 2019 study on resident perceptions of CBD in Québec, which revealed that 74% of participants found CC processes "nebulous" and their decisions "ill-defined," despite having operated under CBD for a year (Fédération des Médecins Résidents du Québec, 2019, p. 7). It is important for programs to liaise with residents to explain the CC decision-making process and ensure that they understand the expectations they must fulfill as learners. In addition to improving engagement, this may also help to increase the amount of data available to CCs, as many programs reported that residents were not initiating assessments as often as they should. Aside from involving residents as CC members where it is appropriate, specific strategies for improving engagement among residents include creating an assessment 'roadmap' that shows where assessment opportunities exist during specific rotations and giving residents regular feedback from CC meetings via a letter or debriefing session with the chair or program director (Kinnear et al., 2018). Another idea may be to invite residents to speak to their performance in front of the committee to more collaboratively identify development opportunities, which could help foster agency and build accountability (Reeve & Shin, 2020). Such a practice would be consistent with other decision-making bodies that evaluate professionals, such as tenure and promotion committees for faculty.

2.4.4 Sharing and Aggregating Data

Given the increasing volume of assessment data to be handled by CCs, robust solutions for sharing and aggregating data are critical. This is especially relevant given the onset of COVID-19, which has resulted in the need for CCs to conduct meetings remotely (see Acai et al., 2018 for guidelines). By the end of the study in July 2020, most programs were transitioning towards a common institutional platform. However, ongoing member training and further development of platform function to meet CCs' needs are required (Thoma et al., 2020). While technology should not be used to replace the in-depth discussion that occurs during CC meetings (Pack et al., 2020), it has the potential to create efficiencies in some areas, such as by pre-flagging residents for review based on certain criteria or allowing data to be aggregated and more easily visualized.

2.4.5 Developing a Clear Mandate

In this study, participants' perspectives of CCs' role in CBD were still evolving. Pack et al. (2020) suggested that in practice, CCs play an important role in supporting resident development and the continuous improvement of educational programs. Most CCs in the present study recognized the need to engage in resident and faculty development; however, there was variability in the extent to which CCs were involved in the development of learning plans. To some extent, this may depend on program size. While in larger programs CCs may exist separately from the RPC, in smaller programs, there may be considerable overlap in the membership of these committees; thus, CCs may perceive a greater responsibility for the development and oversight of learning plans.

Moreover, CCs spent more time discussing residents with performance concerns than residents who were performing well. Prior research suggests that offering development opportunities to all residents, including those performing well, is important (Hauer et al., 2015). Clearly defining CCs' mandate, which may depend on specific contextual factors such as program size, will help committees better enact their charge and ensure that they are appropriately supported in their work.

2.4.6 Limitations and Future Directions

This study was conducted at a single institution, which may limit its generalizability to other centres. Given that sampling was broad (i.e., across multiple programs) and feedback was sought from colleagues at other institutions through the presentation of these data at national and international meetings, the findings of this study likely represent more than just local phenomena. Nonetheless, it is recommended that future research take a multi-site approach that compares and contrasts CC implementation nationally and internationally.

2.5 Conclusions

This study examined CC implementation at a Canadian institution over a three-year period. Overall, CCs were perceived to be a positive addition to residency programs, with most participants believing that they would improve the credibility and defensibility of resident promotion and review processes. Nonetheless, CCs across multiple programs grappled with some challenges as they moved through the various stages of the implementation process, including optimizing membership, maintaining capacity among members, engaging faculty and residents in CBD, sharing and aggregating data, and

developing a clear mandate. Multi-site implementation studies, including those reaching across national boundaries, could help confirm whether these challenges are applicable beyond a single institutional context and promote the sharing of best practices. Future research on CCs should also examine the mechanics of how CCs make decisions in greater detail to better understand the extent to which these committees are able to deliver on their mandate as part of CBME.

2.6 References

Acai, A., Sonnadara, R. R., & O'Neill, T. A. (2018). Getting with the times: A narrative review of the literature on group decision making in virtual environments and implications for promotions committees. *Perspectives on Medical Education, 7*(3), 147-155. https://doi.org/10.1007/s40037-018-0434-9

Andolsek, K., Padmore, J., Hauer, K. E., Edgar, L., & Holmboe, E. (2020). Clinical competency committees: A guidebook for programs (3rd edition). Retrieved from https://www.acgme.org/Portals/0/ACGMEClinicalCompetencyCommitteeGuidebook.pdf

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Chahine, S., Cristancho, S., Padgett, J., & Lingard, L. (2017). How do small groups make decisions? A theoretical framework to inform the implementation and study of clinical competency committees. *Perspectives on Medical Education, 6*(3), 192-198. https://doi.org/10.1007/s40037-017-0357-x

Creuss, S. R., & Creuss, R. L. (2000). Professionalism: A contract between medicine and society. *CMAJ: Canadian Medical Association Journal, 162*(5), 668-669.

Dickey, C. C., Thomas, C., Feroze, U., Nakshabandi, F., & Cannon, B. (2017). Cognitive demands and bias: Challenges facing clinical competency committees. *Journal of Graduate Medical Education, 9*(2), 162-164. https://doi.org/10.4300/JGME-D-16-00411.1

Donato, A. A., Alweis, R., & Wenderoth, S. (2016). Design of a clinical competency committee to maximize formative feedback. *Journal of Community Hospital Internal Medicine Perspectives, 6*(6). https://doi.org/10.3402/jchimp.v6.33533

Doty, C. I., Roppolo, L. P., Asher, S., Seamon, J. P., Bhat, R., Taft, S., . . . Willis, J. (2015). How do emergency medicine residency programs structure their clinical competency committees? A survey. *Academic Emergency Medicine, 22*(11), 1351-1354. https://doi.org/10.1111/acem.12804

Duitsman, M. E., Fluit, C. R. M. G., van Alfen-van der Velden, J. A. E. M., de Visser, M., ten Kate-Booij, M., Dolmans, D. H. J. M., . . . de Graaf, J. (2019). Design and evaluation of a clinical competency committee. *Perspectives on Medical Education, 8*(1), 1-8. https://doi.org/https://doi.org/10.1007/s40037-018-0490-1

Edgar, L., & Holmboe, E. (2019). International clinical competency committees: Maximizing value for faculty, residents, and the program. *Journal of Graduate Medical Education, 11*, 191-192. https://doi.org/10.4300/JGME-D-19-00413

Ekpenyong, A., Baker, E., Harris, I., Tekian, A., Abrams, R., Reddy, S., & Park, Y. S. (2017). How do clinical competency committees use different sources of data to

assess residents' performance on the internal medicine milestones? A mixed methods pilot study. *Medical Teacher, 39*(10), 1074-1083. https://doi.org/10.1080/0142159X.2017.1353070

Fédération des Médecins Résidents du Québec. (2019). Implementation of Competence by Design in Quebec—Year 2: Ongoing issues. Retrieved from http://www.fmrq.qc.ca/files/documents/2f/33/fmrq-report-implementation-of-cbd-in-quebec-year-2-final-english-version-08-2019.pdf

French, J. C., Dannefer, E. F., & Colbert, C. Y. (2014). A systematic approach toward building a fully operational clinical competency committee. *Journal of Surgical Education, 71*(6), e22-27. https://doi.org/10.1016/j.jsurg.2014.04.005

General Medical Council. (n.d.). Annual review of competency progression (ARCP). Retrieved from https://www.gmc-uk.org/education/reports-and-reviews/progression-reports/annual-review-of-competency-progression

Hauer, K. E., Chesluk, B., Iobst, W., Baron, R. B., Boscardin, C. K., ten Cate, O., & O'Sullivan, P. S. (2015). Reviewing residents' competence: A qualitative study of the role of clinical competency committees in performance assessment. *Academic Medicine, 90*(8), 1084-1092. https://doi.org/10.1097/ACM.0000000000000736

Hauer, K. E., ten Cate, O., Boscardin, C. K., Iobst, W., Holmboe, E. S., Chesluk, B., . . . O'Sullivan, P. S. (2016). Ensuring resident competence: A narrative review of the literature on group decision making to inform the work of clinical competency committees. *Journal of Graduate Medical Education, 8*(2), 156-164. https://doi.org/10.4300/JGME-D-15-00144.1

Hemmer, P. A., & Kelly, W. F. (2017). We need to talk: Clinical competency committees in the key of c(onversation). *Perspectives on Medical Education, 6*(3), 141-143. https://doi.org/10.1007/s40037-017-0360-2

Ketteler, E. R., Auyang, E. D., Beard, K. E., McBride, E. L., McKee, R., Russell, J. C., . . . Nelson, M. T. (2014). Competency champions in the clinical competency committee: A successful strategy to implement milestone evaluations and competency coaching. *Journal of Surgical Education, 71*(1), 36-38. https://doi.org/10.1016/j.jsurg.2013.09.012

Kinnear, B., Warm, E. J., & Hauer, K. E. (2018). Twelve tips to maximize the value of a clinical competency committee in postgraduate medical education. *Medical Teacher, 40*(11), 1110-1115. https://doi.org/10.1080/0142159X.2018.1474191

Li, S.-A., Acai, A., Sherbino, J., & Chan, T. M. (in press). The teacher, the assessor, and the patient protector: A conceptual model describing how context interfaces with the supervisory roles of academic emergency physicians. *AEM Education and Training*. https://doi.org/10.1002/aet2.10431

Loo, L. K., Lee, S., & Acosta, D. (2017). Maintaining the public trust in clinical competency committees—societial representatives. *Journal of Graduate Medical Education, 9*(1), 131-132. https://doi.org/10.4300/JGME-D-16-00533.1

Mack, N., Woodsong, C., MacQueen, K. M., Guest, G., & Namey, E. (2005). *Qualitative research methods: a data collector's field guide*. North Carolina: Family Health International.

McQueen, S. A., Petrisor, B., Bhandari, M., Fahim, C., McKinnon, V., & Sonnadara, R. R. (2016). Examining the barriers to meaningful assessment and feedback in medical training. *American Journal of Surgery, 211*(2), 464-475. https://doi.org/10.1016/j.amjsurg.2015.10.002

Monrad, S. U., Mangrulkar, R. S., Woolliscroft, J. O., Daniel, M. M., Hartley, S. E., Gay, T. L., . . . Santen, S. A. (2019). Competency committees in undergraduate medical education: Approaching tensions using a polarity management framework. *Academic Medicine, 94*(12), 1865-1872. https://doi.org/https://doi.org/10.1097/ACM.0000000000002816

Neergaard, M. A., Olesen, F., Anderson, R. S., & Sondergaard, J. (2009). Qualitative description - the poor cousin of health research? *BMC Medical Research Methodology, 9*, 52. https://doi.org/10.1186/1471-2288-9-52

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1), 1-13. https://doi.org/10.1177/1609406917733847

Odorizzi, S., Cheung, W. J., Sherbino, J., Lee, A. C., Thurgur, L., & Frank, J. R. (2020). A signal through the noise: Do professionalism concerns impact the decision-making of competence committees? *Academic Medicine, 95*(6), 896-901. https://doi.org/10.1097/ACM.0000000000003005

Pack, R., Lingard, L., Watling, C. J., Chahine, S., & Cristancho, S. M. (2019). Some assembly required: Tracing the interpretive work of clinical competency

committees. *Medical Education, 53*(7), 723-734. https://doi.org/10.1111/medu.13884

Pack, R., Lingard, L., Watling, C. J., & Cristancho, S. (2020). Beyond summative decision-making: Illuminating the broader roles of competence committees. *Medical Education, 54*(6), 517-527. https://doi.org/10.1111/MEDU.14072

Reeve, J., & Shin, S. H. (2020). How teachers can support students' agentic engagement. *Theory Into Practice, 59*(2), 150-161. https://doi.org/10.1080/00405841.2019.1702451

Royal College of Physicians and Surgeons of Canada. (2014). Competence by design: Reshaping Canadian medical education. Retrieved from http://www.royalcollege.ca/rcsite/documents/educational-strategy-accreditation/royal-college-competency-by-design-ebook-e.pdf

Royal College of Physicians and Surgeons of Canada. (n.d.-a). Competence by Design launch schedule. Retrieved from http://www.royalcollege.ca/rcsite/cbd/cbd-implementation-e

Royal College of Physicians and Surgeons of Canada. (n.d.-b). Competence committees. Retrieved from http://www.royalcollege.ca/rcsite/cbd/assessment/competence-committees-e

Royal College of Physicians and Surgeons of Canada. (n.d.-c). Competence committees—Guidelines for the terms of reference. Retrieved from http://www.royalcollege.ca/rcsite/cbd/assessment/committees/managing-competence-committees-e

Schumacher, D. J., King, B., Barnes, M. M., Elliott, S. P., Gibbs, K., McGreevy, J. F., . . . Appd Learn CCC Study Group. (2018a). Influence of clinical competency committee review process on summative resident assessment decisions. *Journal of Graduate Medical Education, 10*(4), 429-437. https://doi.org/10.4300/JGME-D-17-00762.1

Schumacher, D. J., Martini, A., Bartlett, K. W., King, B., Calaman, S., Garfunkel, L. C., . . . Appd Learn CCC Study Group. (2019a). Key factors in clinical competency committee members' decisions regarding residents' readiness to serve as supervisors: A national study. *Academic Medicine, 94*(2), 251-258. https://doi.org/10.1097/ACM.0000000000002469

Schumacher, D. J., Michelson, C., Poynter, S., Barnes, M. M., Li, S.-T. T., Burman, N., . . . Appd Learn CCC Study Group. (2018b). Thresholds and interpretations: How clinical competency committees identify pediatric residents with performance concerns. *Medical Teacher, 40*(1), 70-79. https://doi.org/10.1080/0142159X.2017.1394576

Schumacher, D. J., Poynter, S., Burman, N., Elliott, S. P., Barnes, M., Gellin, C., . . . Appd Learn CCC Study Group. (2019b). Justifications for discrepancies between competency committee and program director recommended resident supervisory roles. *Academic Pediatrics, 19*(5), 561-565. https://doi.org/10.1016/j.acap.2018.12.003

Thoma, B., Bandi, V., Carey, R., Mondal, D., Woods, R., Martin, L., & Chan, T. (2020). Developing a dashboard to meet competence committee needs: A design-based

research project. *Canadian Medical Education Journal, 11*(1), e16-e34. https://doi.org/https://orcid.org/0000-0003-1124-5786

Tweed, M., & Wilkinson, T. (2019). Student progress decision-making in programmatic assessment: Can we extrapolate from clinical decision-making and jury decision-making? *BMC Medical Education, 19*. https://doi.org/https://doi.org/10.1186/s12909-019-1583-1

Tables and Figures

Table 1. Key findings and corresponding quotes from 24 interviews conducted in Phase 1: Pre-Implementation

Finding	Quotes
Belief that CCs would improve existing promotion and review processes	*"It's a structured format … and there's ownership to these types of decisions."* (Resident) *"It will help with faculty not having to make those decisions at the end of our rotation, which never made sense anyways."* (Faculty)
Divergent views on whether residents should be included as members	*"Their perspective in terms of why somebody is struggling is a valuable add."* (Resident) *"If there is … contention around the decisions that the committee makes, then the resident could be placed in a difficult position in terms of having to defend the committee's decision to their peers."* (Faculty)
Divergent views on external membership: PhD-trained researchers were well-supported but community members outside of the medical education community (e.g., patients) were not	*"We believe that [PhD-trained researchers] will hold the specialty to account, particularly around issues of professionalism."* (Faculty) *"People outside of medicine don't know what it takes to make a [doctor]."* (Resident)
Concern that CCs would increase faculty workload	*"How much effort and time and workload can we displace onto faculty? There's a limitation there, right?"* (Faculty)
Data sharing and aggregation perceived as major challenges to CC implementation	*"So far, our biggest roadblock has been getting information out to people … it's an operational problem that we need to sort out."* (Faculty) *"It's just too much data. If there's anything outstanding, then we bring that up at the committee meeting. But anyone who's passed, there's no point. You've got 60 residents to go through in an hour."* (Faculty)
Lack of clarity about CCs' mandate and how it was to interface with other residency committees	*"Exactly what is meant by a CC is different depending on which publication you read."* (Faculty) *"I don't know exactly what the roles for the CCs within an individual program will be and how they differ from a residency program committee."* (Resident)

Lack of understanding about how CCs make decisions; no best practices to drawn upon	*"Training [around decision-making procedures] is unfortunately happening on the fly, because there is no current, at least in the Canadian context, established functioning CC."* (Faculty) *"We're very aware that our evaluation processes are going to change significantly, but exactly what's going on within the CC is still very unclear."* (Resident)
Program-specific challenges	*"In a smaller program is that the number of voices is limited and it's hard to find the right mix."* (Resident) *"I wouldn't suggest that larger programs have less of a challenge. … You have to cater to 90 to 100 residents and yet you don't have a proportional increase in faculty that are going to be involved."* (Faculty)

Table 2. Key findings and corresponding quotes from 12 interviews conducted in Phase 3: Late Implementation (Two Year Mark)

Finding	**Quotes**
Meetings became more efficient over time	*"What's been positive is that [meetings] have now developed quite a flow to them."* (CC Chair)
Divergent views on the extent to which CCs reduced member workload	"*A lot of the reports have involved collating information and that is my least favourite part because we're all very busy. So, for me to take three hours to go between two systems and input stuff on a spreadsheet … I'm just not sure that that should be the role of the physician committee member."* (CC Member—Faculty) *"In the spectrum of things that academic faculty are expected to do, it's definitely not onerous."* (CC Member—Program Director)
Careful selection of members created a collegial atmosphere in which members felt comfortable sharing their opinions	*"We were very lucky that we chose the right people to be on the committee."* (CC Member—Resident) *"The people that I'm working with are very nice and open-minded. … I feel like I can speak up when it's needed."* (CC Member—Junior Faculty)
Resident representation on the committee brought an important perspective and was a positive experience for residents	*"[Resident representation on the committee] has been working very well. They bring a unique perspective of the learner's view."* (CC Chair) *"I was quite grateful to have been on it, not just from having the personal experience of seeing what it was like, but I also felt quite reassured with the process and the content that was discussed."* (CC Member—Resident)
PhD-trained researcher representation on the committee helped ensure that the committee was applying evaluation criteria consistently and remained accountable for its actions	*"It was beneficial to make sure that we were very clear on the criteria."* (CC Member—PhD-Trained Researcher)
Succession planning became important at later stages of implementation; including faculty from other programs was helpful for building capacity	*"The challenge is going to be finding people to do work that's so labour-intensive."* (CC Member—Faculty) *"It's been interesting for me, joining that committee in advance of me starting as chair of [another program's] CC, to get a sense of how it runs."* (CC Member—Faculty)

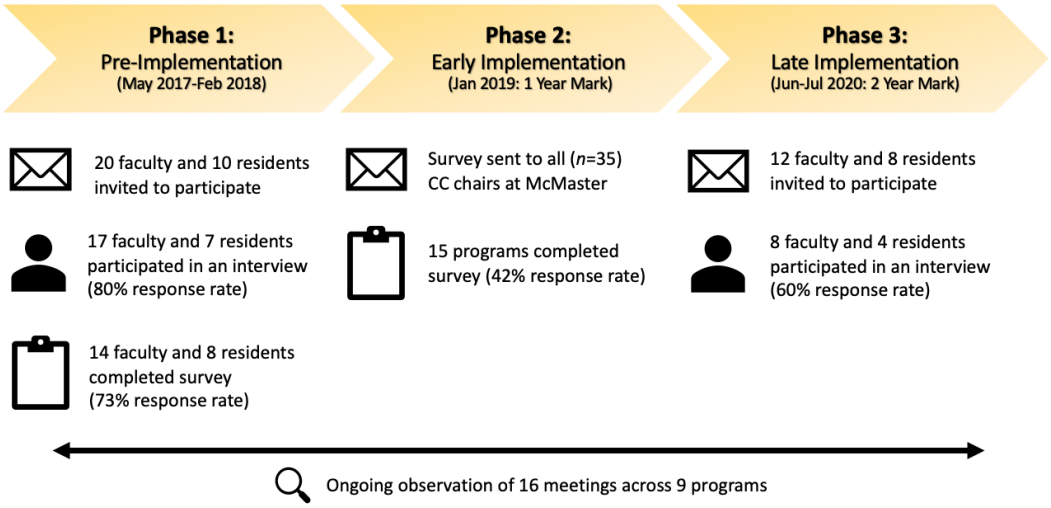

Figure 1. Schematic of the longitudinal study

Chapter 3

Understanding How Individual Raters Make Promotion Decisions

3.1 Introduction

Understanding promotion decisions as category judgments in which a resident is either promoted, conditionally promoted, or not promoted opens the door to a large body of knowledge that can be used to help researchers understand how CCs make decisions. Although CCs are a newer concept in medical education, how physicians make diagnostic decisions—which has also been conceptualized as a categorization task—has been a primary research focus for decades (Elstein, Schulman, & Sprafka, 1990; Monteiro & Norman, 2013). While physicians diagnose patients, CCs 'diagnose' learners, a necessary step in helping them to meet their educational goals. Thus, it is reasonable to suggest that findings from the diagnostic reasoning literature might also apply to CCs. The remaining sections of the introduction will review central concepts from the diagnostic reasoning literature that may be useful for understanding how CCs make decisions. These concepts will be revisited in the discussion, which will explore how they can be used to explain how individual raters make promotion decisions involving anecdotal evidence.

3.1.1 Dual Process Model of Diagnostic Reasoning

The dual process model suggests that physicians use two distinct, yet interrelated cognitive processes when making diagnostic decisions (Croskerry, 2009). The first of these is known as non-analytic or 'System 1' thinking, and is a quick, unconscious, and automatic form of reasoning (Croskerry, 2009). This is common in medicine and occurs when physicians are able to identify the cardinal features, or patterns, of a disease. As an

example, most physicians would be able to recognize a patient with heart palpitations, bulging of the eyes, fine resting tremor, and an enlarged thyroid as presenting with the classic signs of Graves' disease (ten Cate & Durning, 2018). This is in contrast to analytic, or 'System 2' thinking, which is a slow, conscious, and deliberate form of reasoning, such as a physician systematically working through a patient's acid-base status to evaluate their condition (Croskerry, 2009; ten Cate & Durning, 2018).

Analytic and non-analytic reasoning are not mutually exclusive, nor is one necessarily superior to the other (Custers, 2013; Monteiro & Norman, 2013). The need for some automaticity is inevitable, as performing every task in a step-by-step fashion would overload working memory and lead to inefficiency (ten Cate & Durning, 2018; Young, Van Merrienboer, Durning, & ten Cate, 2014). However, there are also tasks in which it is not possible to rely solely on non-analytic reasoning, such as when a physician encounters a case that they do not recognize, or when diagnosis is highly dependent on a set of predefined criteria, as is often the case in psychiatry (Custers, 2013). Combined models of clinical reasoning have also been proposed, which more clearly highlight how clinicians switch between non-analytic and analytic reasoning during a single diagnostic task. For example, Eva (2005) suggested that physicians use non-analytic reasoning to form an initial mental representation of a case, which they then test using more analytic methods, such as history taking, physical examination, and diagnostic testing. This process is bidirectional and reminiscent of Bayesian inference, such that a physician's mental representation and interpretation of a patient's symptoms are refined as different

hypotheses are tested (Etz & Vandekerckhove, 2018; Eva, 2005). In other words, the likelihood of a particular outcome changes as more information becomes available.

Although both experts and novices likely use a combined approach when making diagnostic decisions, there are some notable differences in their clinical reasoning abilities (Eva, 2005; Norman, Young, & Brooks, 2007). The ability to engage in non-analytic reasoning depends on the number of experiences that an individual has had; thus, experts are often faster and more accurate at generating initial hypotheses about a patient's condition than novices (Norman, Rosenthal, Brooks, Allen, & Muzzin, 1989). Experts may also be less likely to encounter cases that they do not recognize; thus, they may not need to rely on analytic reasoning as often as novices (Custers, 2013). Another hallmark of expertise may be the ability to recognize when a change in approach is needed, allowing one to more seamlessly toggle between non-analytical and analytical forms of reasoning (Moulton, Regehr, Mylopoulos, & MacRae, 2007).

3.1.2 The Role of Memory in Diagnostic Reasoning

While analytic and non-analytic reasoning help to explain diagnostic reasoning at a high level, researchers have also attempted to understand the memory processes that are invoked when clinicians engage in these forms of reasoning (Monteiro & Norman, 2013). The basic premise is that a newly encountered stimulus, such as a patient case, triggers a concept in memory that is then used to make category judgments on the basis of similarity (Monteiro & Norman, 2013). Two models have been used to explain how these judgments are made: exemplar (Brooks, 1978; Medin & Schaffer, 1978) and prototype (Rosch & Mervis, 1975). According to the exemplar model, all previously encountered

instances of a category, known as exemplars, are stored in memory (Rouder & Ratcliff, 2006). In medicine, an exemplar could be the "signs and symptoms associated with each individual, previously experienced case" (Papa & Li, 2015, p. 218). When a new stimulus is encountered, it is compared against all previously encountered exemplars to determine which category it most resembles (Rouder & Ratcliff, 2006). This is thought to occur during non-analytic reasoning, which is based on the retrieval of prior experiences stored in memory (Monteiro & Norman, 2013; Papa & Li, 2015).

According to prototype theory, individuals do not store each instance of a category that they have encountered in memory (Rouder & Ratcliff, 2006). Instead, their most common features are abstracted to create a prototype, which is thought to be an 'average' model of a category (Smith & Minda, 2002). In medicine, this might be described as "'what disease 'X' looks like,'" although with experience, prototypes can be refined to include information about how often certain symptoms are associated with a disease (Papa & Li, 2015, p. 218). When a new stimulus is encountered, it is compared against the prototype rather than individual exemplars; however, category judgments are still based on similarity (Rouder & Ratcliff, 2006). It is possible that prototypes are invoked during analytic reasoning, which involves applying specific criteria, or rules, that relate specific features to categories (Monteiro & Norman, 2013).

Which model best explains category judgments has been the subject of intense debate in the literature (Minda & Smith, 2001); however, scientists have concluded that both have explanatory value (Malt, 1989; Rouder & Ratcliff, 2006). For example, the prototype model may better explain categorization in novices, who have limited practical

experience to draw upon when making judgments (Monteiro & Norman, 2013; Norman et al., 2007). Despite their lack of experience, medical students can still use learned criteria from textbooks or lectures to diagnose cases they have not seen before (Papa & Li, 2015). However, this effect breaks down when they are presented with atypical cases whose features do not resemble the average model stored in memory (Papa & Li, 2015). As individuals accumulate experience, they may begin to learn features that help them to tell different stimuli apart (Rouder & Ratcliff, 2006). These features can then be stored as exemplars and used to categorize new stimuli without needing to rely on learned criteria (Rouder & Ratcliff, 2006). This may help explain why encountering even a single, prior exemplar can increase diagnostic accuracy among experienced physicians (Brooks, Norman, & Allen, 1991). It also offers an explanation for an important finding in the diagnosing reasoning literature: context specificity (ten Cate & Durning, 2018).

Context specificity refers to the fact that although two patients may present with identical features, a physician would not necessarily arrive at the same diagnosis for both (ten Cate & Durning, 2018). Aside from the classical signs and symptoms of a disease, physicians also use contextual factors such as a patient's sex, age, and risk factors when making diagnoses (Hobus, Schmidt, Boshuizen, & Patel, 1987). This is captured in a model of clinical reasoning by Durning and Artino (2011), which suggests that diagnostic outcomes are driven by an interaction of factors related to the physician, the patient, and the system. The idea of situated cognition is parsimonious with this model, stressing that knowledge cannot be interpreted outside of the social, cultural, and physical contexts from which it originates (Brown, Collins, & Duguid, 1989; ten Cate & Durning, 2018).

Returning to the idea of prototypes and exemplars, a strictly prototype-based model may not be sensitive enough to explain the level of specificity that has been observed in physicians' diagnoses, as individual features of a case would be averaged out within a prototype (Smith & Minda, 2002). Thus, while prototypes may be helpful for categorizing information when experience is limited, exemplars are likely better at accounting for the individual features of a case (Rouder & Ratcliff, 2006).

3.1.3 The Role of Anecdotal Evidence in Competence Committee Decision-Making Processes

In interviews conducted with CC members as part of this thesis, participants reported valuing 'objective' data sources such as Entrustable Professional Activities (EPAs), yet they were frequently observed to be relying on subjective anecdotes when making decisions. These observations are supported by the literature, which has shown anecdotal evidence to be prevalent in CC decision-making processes (Ekpenyong et al., 2017; Hauer et al., 2015; Pack, Lingard, Watling, Chahine, & Cristancho, 2019; Schumacher et al., 2018). An interview-based study by Hauer and colleagues (2015) found that it was common for CCC members to use patient or nurse complaints, hallway conversations, and emails sent to program leaders when making summative decisions. Similarly, a mixed-methods study by Ekpenyong and colleagues (2017) found that CCC members' personal experiences with residents were among the top three data sources used when making promotion decisions. These findings are supported by the broader literature, which suggests that anecdotes are salient in inferential tasks from individual health decisions to policy making (Enkin & Jadad, 1998; McDonough, 2001). There are likely evolutionary reasons for this, as before humans could write, they had were forced to rely

on personal experiences and stories from trusted individuals when making decisions (Boyd, 2017). Personal experiences are particularly compelling because they are often vivid and generate a strong emotional response, which increases the likelihood that they will be remembered and considered in future decision-making tasks (Nisbett, 1980; Tversky & Kahneman, 1974). Taken together, these findings suggest that anecdotal evidence likely plays an important role in CC decision-making processes; yet, how and why it is used is not yet well understood (Schumacher et al., 2018).

The present chapter begins by investigating how individual raters make promotion decisions involving anecdotal evidence. Studying individual raters is important for at least two reasons. Before CCs were implemented, promotion decisions were made by individual raters, such as longitudinal clinical supervisors or program directors (Royal College of Physicians and Surgeons of Canada, n.d.). In order to understand whether CCs confer any advantages over individual raters, a central premise of competency-based assessment models (Holmboe, Sherbino, Long, Swing, & Frank, 2010), it is first necessary to understand how individual raters make promotion decisions. Studying individual raters is also important for understanding group behaviours, as it is their initial preferences that form the basis for subsequent group decisions (Stasser, 1999).

3.1.4 Study Purpose and Overview

Research on diagnostic reasoning and categorization offers a rich theoretical framework for the research presented in this chapter, which sought to examine how individual raters make promotion decisions. This study consisted of three experiments. Experiment 1 used a sample of novice raters to examine how individual raters respond to

different types of anecdotal evidence when making promotion decisions using simulated resident files. Experiment 2 explored the same question using a sample of experienced raters. After finding that individual raters' promotion decisions deviated significantly from the promotion criterion when anecdotal evidence was provided, Experiment 3 sought to explore this phenomenon in further detail by replicating Experiment 1, but this time asking raters to provide written explanations for each of their promotion decisions.

3.2 Methods

3.2.1 Experiment 1: Novice Raters

Participants in this study were undergraduate students enrolled in the psychology participant pool at McMaster University in Hamilton, Ontario, Canada. The rationale for using undergraduate psychology students as novice raters was twofold. First, capturing the experiences of novice raters was felt to be important for developing training materials for future CC members. However, a practical challenge was that early CC members were typically individuals with prior experience making promotion decisions through processes preceding CCs, and as such, could not be considered true novices. Thus, it was necessary to come up with an alternative way of capturing the perspectives of novices. Second, recruiting busy clinicians to participate in research on CCs is challenging. Thus, this study also aimed to explore whether undergraduate psychology students, who are relatively easy to recruit for the purposes of medical education research, could reasonably be used as a proof-of-concept before recruiting clinicians to participate.

During recruitment, participants were screened to ensure that they did not have any experience making promotion decisions prior to taking part in the study. Recruitment

was done through SONA, an online system that allows students to complete research studies in exchange for course credit. Participants received one course credit in exchange for completing this study. An *a priori* sample size calculation was used as a guideline for participant recruitment. Using G*Power v. 3.1, it was determined that approximately 60 participants would be needed to attain 80% power.

3.2.1.1 Materials

Forty-two simulated resident files were developed by the primary investigator for use in this study. Each file included a resident name and performance data in the form of a ratio of successful: unsuccessful attempts on an EPA, which ranged from competence demonstrated successfully zero out of six times (0:6) to competence demonstrated successfully six out of six times (6:6). EPAs represent the essential skills required of residents in a given specialty and, while not the sole source of data used by CCs, are an important form of assessment data in CBME (ten Cate, 2013). Seven resident files were control cases in which no other information was provided besides performance data, while the other 35 cases contained performance data and one of five written vignettes. The vignettes, shown in Table 1, were designed to simulate anecdotal evidence that a CC member might commonly encounter based on findings from the primary investigator's observations of these committees. The vignettes were intentionally designed to encompass issues relevant to CCs, such as a lack of professionalism, which has been shown to be more predictive of future problems in medicine (i.e., malpractice claims) than a lack of technical skills (Levinson, Roter, Mullooly, Dull, & Frankel, 1997).

In order to increase the study's realism, each resident file included a name randomly drawn from an online database of common English names (http://random-name-generator.info/). An equal number of women's and men's names were used. Prior to the experiments outlined in this chapter, a pilot experiment involving undergraduate students was conducted to ensure that the perceived gender of the residents being evaluated did not unduly influence participants' promotion decisions. Thirty participants (25 females and five males) were randomly assigned to either a group in which they were presented with resident files containing only women's names or a group in which they were presented files containing only men's names. The remainder of the pilot experiment followed the data collection and analysis procedures detailed below, with the findings showing no significant differences in participants' promotion decisions for files with women's names as compared with files with men's names.

3.2.1.2 Procedures

This study was conducted online using LimeSurvey. After consent, participants provided demographic information such as their age, program, and year of study. They were then provided study instructions and definitions of key terms such as "residency," "resident," "CC," and "EPA." Participants were asked to answer all questions while imagining themselves as physician members of a CC. Additionally, to simulate how CCs operate in real life, they were given a minimum passing criterion for a resident to be promoted: the successful completion of an EPA at least three times.

Once participants had reviewed the instructions, they were presented with the 42 simulated resident files described earlier in random order. After reviewing each file, they

were asked to indicate whether they would promote each resident by selecting "Yes" or "No." Optional text boxes were provided with each scenario in case participants wished to share any qualitative comments with the primary investigator.

3.2.1.3 Data Analysis

Demographic data were analyzed using descriptive statistics, such as means, standard deviations, proportions, and percentages. In each condition, the percentage of participants choosing to promote a resident at each level of performance (i.e., 0:6 to 6:0) was calculated by dividing the number of "Yes" responses by the total number of participants and multiplying it by 100. Each experimental condition was then compared with the control using a permutation test to look for any overall differences. Permutation tests allow for the sampling distribution of any test statistic to be computed, under the null hypothesis that a variable of interest has no effect on the outcome (Collingridge, 2013; Rice & Lumley, 2008). Permutation tests are increasingly common in scientific studies, as they more powerful than traditional statistical methods such as *t*-tests and their non-parametric analogs, particularly when working with small sample sizes or non-normal data, such in Experiment 2 (LaFleur & Greevy, 2009). The general premise of a permutation test is that the data are 'shuffled' (i.e., permuted) by assigning different outcome values to each observation from among the set of actually observed outcomes (Collingridge, 2013). This is used to test the null hypothesis that the two samples are drawn from the same population. If the null hypothesis is true, then the shuffled data should look like the experimental data; however, if it is false, then the shuffled data would be expected to look different from the experimental data (Rice & Lumley, 2008).

In this study, permutation tests involved statistically comparing a distribution of sigmoid curves fitted to each experimental condition with the distribution of curves fitted to the control condition. First, the seven data points within each condition were plotted and sigmoid curves were fitted to each using a leave-one-out approach. In other words, an algorithm iterated through the data points, dropping one at a time and fitting a sigmoid curve to the remaining six data points. The purpose of this step was to generate a distribution of possible curves that could fit the data. Next, the growth rate of each curve was extracted and used for permutation testing.

The curve-fitting procedure generated seven parameters for the control condition and each of the experimental conditions. These data points were labelled as belonging to either the control condition or an experimental condition. Permutation involved randomly shuffling these labels, such that there was now a new set of values labelled "control" and a new set labelled "experimental." This shuffling procedure was repeated 100,000 times, which is sufficiently large to create a distribution of mean differences similar to the exact distribution according to guidelines in the literature (Collingridge, 2013). A permutation test statistic (i.e., the difference in sample means) was calculated for each iteration and compared with the observed test statistic (i.e., the difference in sample means of the original samples). A *p*-value was calculated by dividing the number of times the permutation test statistic was equal to or more extreme than the observed test statistic by the total number of permutations. Once *p*-values had been determined for each experimental condition, they were corrected for five comparisons using the Holm-Bonferroni method (Maxwell & Delaney, 2004).

After determining if there were any overall differences, McNemar's test was used to test for significant differences at the midpoint of each curve. It was hypothesized that if there was an effect of anecdotal evidence, it would be most likely to occur in cases with the highest levels of ambiguity, such as in the 3:3 subcondition (Tversky & Kahneman, 1974). McNemar's test was selected due to the dichotomous nature of the outcome variable and the fact that each participant experienced each condition as part of the study design (McNemar, 1947). Given the already-conservative nature of non-parametric tests, resulting *p*-values were not corrected for multiple comparisons to avoid eroding real effects in the data (Maxwell & Delaney, 2004).

Demographic information was analyzed using Microsoft Excel. Curve fitting and permutation tests were conducted using R version 3.5.2, while non-parametric tests were conducted using IBM® SPSS® Statistical Software version 25.

3.2.2 Experiment 2: Experienced Raters

Participants in this study were CC members with approximately two years of involvement in resident promotion and review processes. CC members were recruited in two phases. In April 2018, the primary investigator attended a CBME-related event organized by McMaster University's Postgraduate Medical Education Office where she shared information about the study and circulated a sign-up sheet. Information about the study and an invitation to participate were emailed to all individuals who provided their contact information. In subsequent months, information about the study and an invitation to participate were also circulated to a number of CC chairs at McMaster, who were asked

to forward the information to their members. Some chairs also forwarded the information to their contacts at other Canadian institutions.

Theoretically, the same sample size calculation used for Experiment 1 also applies to this experiment. However, for feasibility reasons, it was not possible to obtain a sample of 60 experienced raters to participate in the study as this would have exceeded the population of CC members within the institution at the time of the study. Thus, this experiment aimed for a smaller sample of 20 participants to ensure that data collection could be completed in a reasonable period of time.

3.2.2.1 Materials

The same 42 simulated resident files were used as in Experiment 1.

3.2.2.2 Procedures

The same procedures were followed as in Experiment 1, except the demographic questionnaire was revised to gather information about participants' role(s), clinical specialty, and experience on a CC and/or other promotion and review committees. Participants were instructed to make promotion decisions as they would in real life.

3.2.2.3 Data Analysis

The same data analysis procedures were used as in Experiment 1.

3.2.3 Experiment 3: Reasoning

Participants were undergraduate students enrolled in the psychology participant pool at McMaster University in Hamilton, Ontario, Canada. Given that data from this study were analyzed descriptively rather than inferentially, an *a priori* sample size calculation was not needed. Instead, the study was advertised to students via the SONA

research participation system for two weeks, during which any interested students could register for and complete the study in exchange for one course credit.

3.2.3.1 Materials

The same 42 simulated resident files were used as in Studies 1 and 2.

3.2.3.2 Procedures

The same procedures were followed as in Experiment 1, except participants were asked to provide a written reason for each of their promotion decisions.

3.2.3.3 Data Analysis

Demographic information was analyzed as in Experiments 1 and 2. Written explanations of participants reasoning for their promotion decisions were transferred into an electronic spreadsheet and deidentified by removing any information about participants and the resident's performance level (i.e., ratio of successful: unsuccessful EPAs). Analysis of participants' written reasoning was done deductively, through the application of a set of pre-determined codes. Two raters examined each comment independently and placed it into one of four categories derived from prior research investigating the types of reasoning that groups can use to arrive at a final decision (Kaplan & Miller, 1987): 1) rule-based, 2) judgmental, 3) both, or 4) cannot be determined. Rule-based reasoning encompassed situations in which participants were clearly following the rule given to them at the start of the experiment. Judgmental reasoning was defined as reasoning in which the participant appeared to be using their own judgment as opposed to the rule. Cases in which participants used both judgmental and rule-based reasoning were labelled "both." In a small number of cases, participants

misunderstood the instructions and gave responses that did not fit into any of these three categories. In these cases, responses were labelled "cannot be determined."

Once each rater had individually categorized each comment, they met to compare codes. Initially, inter-rater agreement as measured by Cohen's kappa was 0.62, which is considered moderate (McHugh, 2012). At this time, it was determined that one of the raters was attempting to classify comments that would more appropriately have been labelled "cannot be determined," as the participants had clearly misunderstood the premise of the experiment. Once this issue was resolved by ensuring that both raters were clear about the purpose and boundaries of the experiment, the discrepant comments were re-categorized and a Cohen's kappa of 1.00 was attained.

All analyses related to this experiment were conducted using Microsoft Excel.

3.2.4 Ethics Approval

The studies outlined in this chapter were approved by the Hamilton Integrated Research Ethics Board (HiREB-4534).

3.3 Results

3.3.1 Experiment 1: Novice Raters

Sixty undergraduate students (51 females, nine males) participated in this experiment. The majority of participants were in first year (n = 28; 47%), and their most common program of study was life sciences (n = 23; 38%). Participants did not report any prior experience with making promotion decisions prior to participating in the study, confirming that the sample consisted of novice raters.

Figures 1a to b and 1c to e show novice raters' promotion decisions in conditions with positive and negative anecdotal evidence about a resident's performance, respectively. Compared with the control, participants became significantly more lenient in their promotion decisions when presented with positive anecdotes (i.e., strict assessors and prior positive experience) and significantly stricter when presented with negative anecdotes (i.e., lenient assessors, difficult personal circumstance, and unprofessional behaviour). Permutation tests showed that the decision curve associated with each experimental condition differed significantly from the control curve (Table 2). McNemar's tests confirmed that there were significant differences between the experimental conditions and the control at the midpoint of each curve in all conditions except difficult personal circumstance (Table 2). Recall that the midpoint of each curve reflected the point of greatest ambiguity in each condition (i.e., the 3:3 subcondition).

Post-hoc McNemar's tests were conducted to examine the extent of the observed effects. For negative anecdotes, significant effects were observed at all levels of performance above the midpoint (i.e., 4:2 to 6:0), with the exception of the 5:1 subcondition in the lenient assessors condition. It is noteworthy that while a significant effect was not observed at the midpoint of the prior positive experience curve, significant effects were seen at all levels of performance above the midpoint. For positive anecdotes, significant effects were also observed beyond the midpoint, but this time in the opposite direction. These effects did not extend as far as for negative anecdotes, however, as they were observed only in the 2:4 subcondition.

Figures 2a and b show the variability in novice raters' promotion decisions for two representative conditions: prior positive experience (a positive anecdote) and unprofessional behaviour (a negative anecdote), respectively. Of note is the amount of variability present in individual raters' promotion decisions, with the most variability seen among cases with the highest levels of ambiguity in each condition.

3.3.2 Experiment 2: Experienced Raters

Twenty CC members (nine females, 11 males) participated in this experiment, including 18 clinical faculty, one resident, and one non-clinical member. Participants had served on a CC for an average of nine months at the time of the experiment; the minimum length of service was one month. The majority (n = 16; 80%) of participants also reported having at least two years of experience with resident promotion processes outside of a CC, confirming that the sample consisted mostly of experienced raters. Participants represented five academic institutions across Canada and nine specialties, including anesthesiology, emergency medicine, general surgery, medical oncology, nephrology, pathology, pediatrics, obstetrics and gynecology, and orthopedic surgery.

Figures 3a to b and 3c to e show experienced raters' promotion decisions in conditions with positive and negative anecdotal evidence about a resident's performance, respectively. Unlike novice raters, experienced raters did not become more lenient in their promotion decisions as a result of positive anecdotes; however, they did become significantly stricter when presented with negative anecdotes. Permutation tests showed significant differences between the decision curves associated with the lenient assessors and unprofessional behaviour conditions and the control curve (Table 3). McNemar's

tests confirmed that there were significant differences between each of these conditions and the control at the midpoint of each curve (Table 3).

Post-hoc McNemar's tests revealed that only the unprofessional behaviour condition showed an effect beyond the midpoint, with significant effects observed at two performance levels above this point (4:2 and 5:1). Thus, the effect of anecdotal evidence among experienced raters was less pronounced than among novices. There was also less variability among experienced raters' promotion decisions than those of novices (Figure 4a); however, certain anecdotes could still lead to considerable variability among experienced raters, such as those concerning unprofessional behaviour (Figure 4b).

3.3.3 Experiment 3: Reasoning

Thirteen undergraduate students (10 females, three males) participated in this experiment. The majority of participants were in first year (n = 8; 62%), and their most common programs of study were life sciences (n = 6; 46%) and nursing (n = 3; 23%). Participants did not report any prior experience with making promotion decisions prior to participating in the study, confirming that the sample consisted of novice raters.

A total of 515 written explanations were categorized by two raters. Of these, 211 (41%) were classified as rule-based, 100 (19%) as judgmental, and 157 (31%) as both. A classification could not be determined for 47 (9%) of the explanations. Tables 4a to f detail the types of reasoning used by participants in each condition.

Participants tended to use rule-based reasoning (e.g., "The resident did not complete the minimum # of EPAs") when the situation was straightforward, such as when a resident clearly did not meet the formal promotion criteria. In situations where

anecdotal evidence was congruent with the available performance data, participants often combined both rule-based and judgmental reasoning to justify their decisions (e.g., "The resident performed poorly *and* was unprofessional"). The most interesting cases, however, were those in which participants used their judgment to deliberately override formal promotion criteria. In these situations, it became clear that participants were *deliberately* choosing to incorporate anecdotal evidence into their decision making, rather than this happening unconsciously or accidentally. In these cases, participants justified their decisions with reasons such as, "Acting unprofessionally trumps test scores."

3.4 Discussion

This study investigated how individual raters make promotion decisions. Experiment 1, which had novice raters individually make promotion decisions about simulated resident files with and without anecdotal evidence, showed that anecdotal evidence could result in novice raters becoming significantly stricter or more lenient in their promotion decisions, depending on the nature of the evidence. Experiment 2 used the same paradigm with experienced raters and found that while they did not become more lenient in their promotion decisions as a result of anecdotal evidence, they became significantly stricter in light of certain, negative anecdotes (e.g., lenient assessors or unprofessional behaviour). Experiment 3 was a replication of Experiment 1, but this time requesting that raters provide a written explanation for their promotion decisions. Findings revealed that individual raters' decision to use anecdotal evidence when making promotion decisions was deliberate and based on conscious judgment.

3.4.1 The Role of Anecdotal Evidence in Competence Committee Decision-Making Processes

A central finding of this study was individual raters' sensitivity to anecdotal evidence about a resident's performance, as well as the fact that raters deliberately (rather than accidentally or unconsciously) incorporated this information into their decision making. This corroborates the observational data collected as part of this thesis, as well as literature suggesting that humans—including CC members—readily use anecdotal evidence when making decisions, perhaps for evolutionary reasons (Ekpenyong et al., 2017; Etz & Vandekerckhove, 2018; Hauer et al., 2015; McDonough, 2001; Pack et al., 2019; Schumacher et al., 2018). In addition to the salience of anecdotes, situated cognition, or the idea that knowledge cannot exist independently of the context in which it originates, has previously been used to explain why physicians' diagnoses depend on an interconnected web of factors related to the physician, the patient, and the system (Brown et al., 1989; ten Cate & Durning, 2018). This idea may apply to CCs, as well, since it appears that performance data cannot meaningfully be interpreted independently of context, in the same way that lab values cannot meaningfully be interpreted without considering a patient's sex, age, or risk factors (Hobus et al., 1987).

Anecdotal evidence may be especially helpful in situations of high ambiguity. In Experiments 1 and 2, the effect of anecdotal evidence was most pronounced at the midpoints of each curve (i.e., the 3:3 subcondition), which reflected the point of highest ambiguity within each condition. Although CCs may not encounter the same contrived situation in real life, it is common for them to have to grapple with 'problematic' data that are missing, lacking in credibility, or misaligned with their perceptions of a resident (Pack

et al., 2019). In fact, after studying CCs across multiple residency programs, Pack and colleagues concluded that all "data (even seemingly [sic] objective data like standardized test scores) … require human assembly, interpretation and judgment" (p. 731). Anecdotal evidence, while subject to interpretation itself, may be one way for CC members to gain the context needed to appropriately contextualize the data in a resident's file.

Certain types of anecdotal evidence may be more influential than others. In this study, negative anecdotes were more salient than positive anecdotes, particularly when they hinted at unprofessional behaviour. While novice raters were sensitive to positive anecdotes, the effect was not as pronounced as for negative anecdotes; moreover, experienced raters were responsive only to negative anecdotes. This may be explained by negativity bias, the idea that humans are generally more sensitive to negative information than positive information, as it is more likely to signal impending danger (Rozin & Royzman, 2001). Another explanation is that in medicine, the majority of future problems (i.e., malpractice claims) are related to issues of communication and professionalism rather than a lack of technical skills (Levinson et al., 1997). Thus, CC members may be especially sensitive to information that could indicate problems related to a resident's professionalism (Odorizzi et al., 2020; Schumacher et al., 2018). A third possible explanation may be that anecdotes about unprofessionalism relate to personal failings (e.g., dishonesty, lack of emotional intelligence, impulse control), which are more difficult to change than situational factors, such as someone going through a difficult time or not having had sufficient practice on a given skill.

3.4.2 Differences Between Novice and Experienced Raters

Although both novice and experienced raters displayed a sensitivity to anecdotal evidence, there were some notable differences between the two groups. Experienced raters were less sensitive to anecdotal evidence than novices and displayed less variability in their promotion decisions. Clinical context and, more specifically, an awareness of the costs of unprofessionalism (e.g., a higher probability of future malpractice claims) may account for some of these differences, as experienced clinicians may have been more likely than undergraduate psychology students to see themselves as 'gatekeepers of their profession' who, above all else, are responsible for ensuring patient safety (Li, Acai, Sherbino, & Chan, in press). This may have prevented them from promoting residents early, since doing so without documented evidence of a resident's competence could pose a threat to not only patient safety but also to their own liability as CC members. Novice raters, on the other hand, would not have had this context, perhaps leading them to behave more like peer assessors than as gatekeepers of the profession. This may parallel how residents serving on CCs would behave, although further research is needed to confirm this. Manipulating the experimental instructions for novice raters to inform them about the potential costs of unprofessionalism could help test the hypothesis that differences are the result of clinical context, as one would expect a reduction in or elimination of differences between novice and experienced raters if this hypothesis was true.

Differences in the number of prior cases encountered may also have accounted for some of the findings. For most novice raters, these experiments would have been the first time they made promotion decisions. Without any prior experience to draw upon, they

may have been relying on crude prototypes of what good or bad resident performance looks like (Monteiro & Norman, 2013; Norman et al., 2007), perhaps gleaned from the study instructions or their own experiences as patients—e.g., "A good doctor is nice and doesn't make many mistakes." While these prototypes would have allowed them to distinguish between poor and excellent performance, residents whose performance was borderline may have been harder to evaluate. For similar reasons, novice raters may also have been less discerning about anecdotal evidence, as it would have been difficult for them to know when it was meaningful versus irrelevant to a resident's future performance. Experienced raters, on the other hand, had approximately two years of experience making promotion decisions before participating in the study. Exposure to a larger number of prior exemplars may have made them more adept at picking up on features important for distinguishing between residents at different performance levels (Norman et al., 2007). They may have been particularly attuned to features that are warning signs of problems to come, such as residents seen to be 'gaming the system' by selecting only lenient assessors or anecdotes about unprofessionalism from colleagues (Acai, Li, Sherbino, & Chan, 2019; Odorizzi et al., 2020; Schumacher et al., 2018). Experienced raters may also have been more confident in their ability to interpret performance data, making it less necessary for them to rely on anecdotes than novices.

Observational data support the use of exemplars among experienced CC members. During the primary investigator's observations of CCs, members with substantial institutional memory, such as the program director, would sometimes comment on resident files reminded them of "Alice from three years ago who was great" or "Sam from

last year who was problematic." This suggests that they were comparing new files to individual exemplars stored in memory. Recalling resident files seen in the past may be easier to do than recalling specific disease exemplars, since resident files are associated with people likely to be well known to CC members whereas disease exemplars may originate from only brief interactions with patients during a busy clinical day.

At a higher level, Eva's (2005) combined model of clinical reasoning may also be applicable to explaining how CC members make promotion decisions. Given that most residency programs review multiple resident files per meeting, often under tight time constraints, some form of pattern recognition is likely necessary for committees to achieve a basic level of efficiency in their work (Chahine, Cristancho, Padgett, & Lingard, 2017; ten Cate & Durning, 2018; Young et al., 2014). A quick, initial glance at each resident file may enable CC members to generate a hypothesis about how a learner is progressing that can inform whether a more in-depth review of their file is needed. If any flags are identified, members can switch to a more analytic process that allows them to review the file in detail; otherwise, they may spend only a few minutes on it and move on. This process is consistent with Pack and colleagues' (2019) description of CCs moving between 'effortful' and 'effortless' decision-making processes, although they did not explicitly link their findings to the cognition literature. Akin to diagnosticians, the ability to scan through files and correctly recognize flags that signal the need for in-depth review may be a sign of expertise among CC members (Moulton et al., 2007). While experienced members can more seamlessly toggle between non-analytic and analytic forms of reasoning, novice raters' lack of prior experience may cause them to either miss

important details or spend too much time focusing on the wrong details, thereby leading to inefficiency. This is supported by the observation that novices took longer to complete the experiment than experienced raters (28 versus 22 minutes, respectively), potentially signaling a greater reliance on analytic reasoning than on pattern recognition.

3.4.3 Variability Among Individual Raters' Promotion Decisions

Another notable finding was the amount of variability in individual raters' promotion decisions. While novice raters showed more variability in their promotion decisions, experienced raters still exhibited considerable variability under certain conditions, such as when presented with anecdotes about unprofessional behaviour. The practical implication of this finding is in models that rely on individual raters such as clinical teachers or program director to make promotion decisions, one rater may choose to make decisions very differently than another, potentially leading to inconsistent application of the promotion criteria and poor defensibility of promotion decisions.

A potential solution to this challenge is the use of groups to make promotion decisions, which is the premise behind the use of CCs in competency-based assessment models (Holmboe et al., 2010). However, to date, no research has tested the extent to which the purported benefits of using groups to make promotion decisions are actually realized in practice, as much of the existing literature on CCs is theoretical in nature. Thus, there is limited empirical research on which to base training or other interventions that might help CCs make more robust decisions. A central premise of social decision scheme theory, which has been used as a theoretical framework to study groups in a variety of contexts, is that groups tend to 'smooth out' the variability seen among

individual raters (Stasser, 1999). This likely has to do with their information processing capabilities, which enable them to not only share more information, but also to scrutinize that information more thoroughly (Hinsz, Tindale, & Vollrath, 1997). This will be tested in Chapter 4, which uses a similar paradigm to investigate the role that anecdotal evidence plays in decisions made by groups.

3.4.4 Limitations and Future Directions

The findings of this study should be considered in light of several limitations. The use of undergraduate psychology students as novice raters was a practical solution to the fact that most early CC members had prior experience making promotion decisions in other contexts and were therefore not true novices. However, since these students did not have any medical experience, they may have behaved differently than would novice clinicians. As CC membership expands over time, studies should explore the extent to which promotion decisions may differ between novice and experienced clinicians. Furthermore, in designing the stimuli used in this chapter, it was necessary to simplify resident files so that the effect of anecdotal evidence could be isolated. However, this also means that they were not fully representative of the complex data that CCs encounter in real life. This limitation will be addressed in Chapter 4, which considers how CCs make decisions involving anecdotal evidence in naturalistic environments.

3.5 Conclusion

This chapter considered the role that anecdotal evidence plays in novice and experienced raters' individual decision-making processes. Both types of raters were sensitive to certain types of anecdotal evidence and the decision to incorporate it into

raters' promotion decisions was deliberate, suggesting that this information may provide information that is helpful for contextualizing performance data, particularly when these data are ambiguous. Nonetheless, there were some differences in how experienced versus novice raters made promotion decisions involving anecdotal evidence, which may be attributed to experienced raters' superior pattern recognition abilities. Substantial variability among novice raters' promotion decisions, and at times among experienced raters, presents a practical challenge in ensuring that promotion criteria are consistently applied and that promotion decisions are defensible. Chapter 4 will explore the extent to which group decision making can offer a solution to this problem.

3.6 References

Acai, A., Li, S.-A., Sherbino, J., & Chan, T. M. (2019). Attending emergency physicians' perceptions of a programmatic workplace-based assessment system: The McMaster Modular Assessment Program (McMAP). *Teaching and Learning in Medicine, 31*(4), 434-444. https://doi.org/10.1080/10401334.2019.1574581

Boyd, B. (2017). The evolution of stories: From mimesis to language, from fact to fiction. *WIREs Cognitive Science, 9*(1), e1444. https://doi.org/10.1002/wcs.1444

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 169-211). Hillsdale, NJ: Erlbaum.

Brooks, L. R., Norman, G., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120*(3), 278-287. https://doi.org/10.1037//0096-3445.120.3.278

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42. https://doi.org/10.3102/0013189X018001032

Chahine, S., Cristancho, S., Padgett, J., & Lingard, L. (2017). How do small groups make decisions? A theoretical framework to inform the implementation and study of clinical competency committees. *Perspectives on Medical Education, 6*(3), 192-198. https://doi.org/10.1007/s40037-017-0357-x

Collingridge, D. S. (2013). A primer on quantitized data analysis and permutation testing. *Journal of Mixed Methods Research, 7*(1), 81-97. https://doi.org/10.1177/1558689812454457

Croskerry, P. (2009). Clinical cognition and diagnostic error: Applications of a dual process model of reasoning. *Advances in Health Sciences Education, 14*(Suppl 1), 27-35. https://doi.org/10.1007/s10459-009-9182-2

Custers, E. J. F. M. (2013). Medical education and cognitive continuum theory: An alternative perspective on medical problem solving and clinical reasoning. *Academic Medicine, 88*(8), 1074-1080. https://doi.org/10.1097/ACM.0b013e31829a3b10

Durning, S. J., & Artino, A. R. (2011). Situativity theory: A perspective on how participants and the environment can interact: AMEE Guide no. 52. *Medical Teacher, 33*(3), 188-199. https://doi.org/10.3109/0142159X.2011.550965

Ekpenyong, A., Baker, E., Harris, I., Tekian, A., Abrams, R., Reddy, S., & Park, Y. S. (2017). How do clinical competency committees use different sources of data to

assess residents' performance on the internal medicine milestones? A mixed methods pilot study. *Medical Teacher, 39*(10), 1074-1083. https://doi.org/10.1080/0142159X.2017.1353070

Elstein, A. S., Schulman, L. S., & Sprafka, S. A. (1990). Medical problem solving: A ten-year retrospective. *Evaluation and the Health Professions, 13*(1), 5-36. https://doi.org/10.1177/016327879001300102

Enkin, M. W., & Jadad, A. R. (1998). Using anecdotal information ini evidence-based health care: Heresy or necessity? *Annals of Oncology, 9*(9), 963-966. https://doi.org/10.1023/A:1008495101125

Etz, A., & Vandekerckhove, J. (2018). Introduction to Bayesian inference for psychology. *Psychonomic Bulletin & Review, 25*, 5-34. https://doi.org/10.3758/s13423-017-1262-3

Eva, K. W. (2005). What every teacher needs to know about clinical reasoning. *Medical Education, 39*(1), 98-106. https://doi.org/10.1111/j.1365-2929.2004.01972.x

Hauer, K. E., Chesluk, B., Iobst, W., Baron, R. B., Boscardin, C. K., ten Cate, O., & O'Sullivan, P. S. (2015). Reviewing residents' competence: A qualitative study of the role of clinical competency committees in performance assessment. *Academic Medicine, 90*(8), 1084-1092. https://doi.org/10.1097/ACM.0000000000000736

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121*(1), 43-64. https://doi.org/10.1037/0033-2909.121.1.43

Hobus, P. P. M., Schmidt, H. G., Boshuizen, H. P. A., & Patel, V. L. (1987). Contextual factors in the activation of first diagnostic hypotheses: Expert-novice differences. *Medical Education, 21*(6), 471-476. https://doi.org/10.1111/j.1365-2923.1987.tb01405.x

Holmboe, E. S., Sherbino, J., Long, D. M., Swing, S. R., & Frank, J. R. (2010). The role of assessment in competency-based medical education. *Medical Teacher, 32*(8), 676-682. https://doi.org/10.3109/0142159X.2010.500704

Kaplan, M. F., & Miller, C. E. (1987). Group decision making and normative versus informational influence: Effects of type of issue and assigned decision rule. *Journal of Personality and Social Psychology*, *53*(2), 306-313. https://doi.org/10.1037/0022-3514.53.2.306

LaFleur, B. J., & Greevy, R. A. (2009). Introduction to permutation and resampling-based hypothesis tests. *Journal of Clinical Child and Adolescent Psychology, 38*(2), 286-294. https://doi.org/10.1080/15374410902740411

Levinson, W., Roter, D. L., Mullooly, J. P., Dull, V. T., & Frankel, R. M. (1997). Physician-patient communication: The relationship with malpractice claims among primary care physicians and surgeons. *Journal of the American Medical Association, 277*(7), 553-559. https://doi.org/10.1001/jama.1997.03540310051034

Li, S.-A., Acai, A., Sherbino, J., & Chan, T. M. (in press). The teacher, the assessor, and the patient protector: A conceptual model describing how context interfaces with the supervisory roles of academic emergency physicians. *AEM Education and Training*. https://doi.org/10.1002/aet2.10431

Malt, B. C. (1989). An on-line investigation of prototype and exemplar strategies in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(4), 539-555. https://doi.org/10.1037/0278-7393.15.4.539

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data* (2nd ed.). New York: Taylor & Francis.

McDonough, J. E. (2001). Using and misusing anecdote in policy making. *Health Affairs, 20*(1), 207-212. https://doi.org/10.1377/hlthaff.20.1.207

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica (Zagreb), 22*(3), 276-282.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12*(2), 153-157. https://doi.org/10.1007/BF02295996

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*(3), 207-238. https://doi.org/10.1037/0033-295X.85.3.207

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(3), 775-799. https://doi.org/10.1037/0278-7393.27.3.775

Monteiro, S. D., & Norman, G. R. (2013). Diagnostic reasoning: Where we've been, where we're going. *Teaching and Learning in Medicine, 25*(S1), S26-S32. https://doi.org/10.1080/10401334.2013.842911

Moulton, C. E., Regehr, G., Mylopoulos, M., & MacRae, H. M. (2007). Slowing down when you should: A new model of expert judgment. *Academic Medicine, 82*(10), S109-S116. https://doi.org/10.1097/ACM.0b013e3181405a76

Nisbett, R. E. (1980). *Human inference: Strategies and shortcomings in social judgement*. Englewood Cliffs, NJ: Prentice-Hall.

Norman, G. R., Rosenthal, D., Brooks, L. R., Allen, S. W., & Muzzin, L. J. (1989). The development of expertise in dermatology. *Archives of Dermatology, 125*(8), 1063-1068. https://doi.org/10.1001/archderm.1989.01670200039005

Norman, G. R., Young, M. E., & Brooks, L. R. (2007). Non-analytical models of clinical reasoning: The role of experience. *Medical Education, 41*(12), 1140-1145. https://doi.org/10.1111/j.1365-2923.2007.02914.x

Odorizzi, S., Cheung, W. J., Sherbino, J., Lee, A. C., Thurgur, L., & Frank, J. R. (2020). A signal through the noise: Do professionalism concerns impact the decision-making of competence committees? *Academic Medicine, 95*(6), 896-901. https://doi.org/10.1097/ACM.0000000000003005

Pack, R., Lingard, L., Watling, C. J., Chahine, S., & Cristancho, S. M. (2019). Some assembly required: Tracing the interpretive work of clinical competency committees. *Medical Education, 53*(7), 723-734. https://doi.org/10.1111/medu.13884

Papa, F. J., & Li, F. (2015). Evidence of the preferential use of disease prototypes over case exemplars among early year one medical students prior to and following

diagnostic training. *Diagnosis, 2*(4), 217-225. https://doi.org/10.1515/dx-2015-0024

Rice, K., & Lumley, T. (2008). Permutation tests. Retrieved from https://faculty.washington.edu/kenrice/sisg/SISG-08-06.pdf

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573-605. https://doi.org/10.1016/0010-0285(75)90024-9

Rouder, J. N., & Ratcliff, R. (2006). Comparing exemplar- and rule-based theories of categorization. *Current Directions in Psychological Science, 15*(1), 9-13. https://doi.org/10.1111/j.0963-7214.2006.00397.x

Royal College of Physicians and Surgeons of Canada. (n.d.). Competence committees. Retrieved from http://www.royalcollege.ca/rcsite/cbd/assessment/competence-committees-e

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*(4), 296-230. https://doi.org/10.1207/S15327957PSPR0504_2

Schumacher, D. J., Michelson, C., Poynter, S., Barnes, M. M., Li, S.-T. T., Burman, N., . . . Appd Learn CCC Study Group. (2018). Thresholds and interpretations: How clinical competency committees identify pediatric residents with performance concerns. *Medical Teacher, 40*(1), 70-79. https://doi.org/10.1080/0142159X.2017.1394576

Smith, D. J., & Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 800-811. https://doi.org/10.1037//0278-7393.28.4.800

Stasser, G. (1999). A primer of social decision scheme theory: Models of group influence, competitive model-testing, and prospective modeling. *Organizational Behavior and Human Decision Processes, 80*(1), 3-20. https://doi.org/10.1006/obhd.1999.2851

ten Cate, O. (2013). Nuts and bolts of Entrustable Professional Activities. *Journal of Graduate Medical Education, 5*(1), 157-158. https://doi.org/10.4300/JGME-D-12-00380.1

ten Cate, O., & Durning, S. J. (2018). Understanding clinical reasoning from multiple perspectives: A conceptual and theoretical overview. In O. ten Cate, E. J. F. M. Custers, & S. J. Durning (Eds.), *Principles and practices of case-based clinical reasoning education. Innovation and change in professional education* (Vol. 15, pp. 35-46). Cham, Switzerland: Springer.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Young, J. Q., Van Merrienboer, J., Durning, S. J., & ten Cate, O. (2014). Cognitive load theory: Implications for medical education: AMEE Guide No. 86. *Medical Teacher, 36*(5), 371-384. https://doi.org/10.3109/0142159X.2014.889290

Tables and Figures

Table 1. Vignettes featuring various types of anecdotal evidence

Title	Description
Strict assessors *(Positive anecdote)*	"In taking a closer look at the evaluations, you notice that the successful evaluations were completed by staff members known to be strict and the unsuccessful evaluations were completed by staff members known to be lenient."
Prior positive experience *(Positive anecdote)*	"You have worked with this resident before. They have always performed well on this EPA when working with you."
Lenient assessors *(Negative anecdote)*	"In taking a closer look at the evaluations, you notice that the successful evaluations were completed by staff members known to be lenient and the unsuccessful evaluations were completed by staff members known to be strict."
Difficult personal circumstance *(Negative anecdote)*	"This resident recently came to you to seek support. They have been going through a difficult personal circumstance—the death of a close family member. They have been finding it difficult to focus on their clinical duties."
Unprofessional behaviour *(Negative anecdote)*	"You were recently approached in the hallway by a fellow staff member you respect, who let you know that this resident had recently acted very unprofessionally with a patient. This staff member has not completed any EPA evaluations for this resident but expressed serious concern about their performance."

Table 2. Novice raters: Comparison of each experimental condition with the control condition using permutation tests to compare entire curves and McNemar's test to compare curve midpoints (i.e., the 3:3 subcondition)

Comparison	**Permutation Test (entire curves)**	**McNemar's Test (at curve midpoints)**
Strict assessors vs. control *(Positive anecdote)*	$p < .01$*	$p = .04$*
Prior positive experience vs. control *(Positive anecdote)*	$p < .01$*	$p < .01$*
Lenient assessors vs. control *(Negative anecdote)*	$p < .01$*	$p = .03$*
Difficult personal circumstance vs. control *(Negative anecdote)*	$p < .01$*	$p = .56$
Unprofessional behaviour vs. control *(Negative anecdote)*	$p < .01$*	$p < .01$*

*Significant at $p < .05$

Table 3. Experienced raters: Comparison of each experimental condition with the control condition using permutation tests to compare entire curves and McNemar's test to compare curve midpoints (i.e., the 3:3 subcondition)

Comparison	**Permutation Test (entire curves)**	**McNemar's Test (at curve midpoints)**
Strict assessors vs. control *(Positive anecdote)*	$p = .13$	$p = 1.00$
Prior positive experience vs. control *(Positive anecdote)*	$p = .13$	$p = 1.00$
Lenient assessors vs. control *(Negative anecdote)*	$p = .01$*	$p = .02$*
Difficult personal circumstance vs. control *(Negative anecdote)*	$p = .06$	$p = .50$
Unprofessional behaviour vs. control *(Negative anecdote)*	$p < .01$*	$p < .01$*

*Significant at $p < .05$

Table 4a. Control condition: Number and percentage of novice raters choosing to promote or not promote residents at various levels of performance, along with the predominant forms of reasoning used to justify their promotion decisions

# Successful EPAs	*n* (%) Promoting	Predominant Reasoning	*n* (%) Not Promoting	Predominant Reasoning
0-2	0 (0%)	--	39 (100%)	Rule-based
3	10 (71%)	Rule-based	4 (29%)	X
4-6	38 (97%)	Rule-based	1 (3%)	Judgmental

X = cannot be classified

Table 4b. Strict assessors condition: Number and percentage of novice raters choosing to promote or not promote residents at various levels of performance, along with the predominant forms of reasoning used to justify their promotion decisions

# Successful EPAs	*n* (%) Promoting	Predominant Reasoning	*n* (%) Not Promoting	Predominant Reasoning
0-2	1 (3%)	Judgmental	34 (97%)	Rule-based
3	11 (92%)	Both	1 (8%)	Judgmental
4-6	36 (100%)	Both	--	--

Table 4c. Prior positive experience condition: Number and percentage of novice raters choosing to promote or not promote residents at various levels of performance, along with the predominant forms of reasoning used to justify their promotion decisions

# Successful EPAs	*n* (%) Promoting	Predominant Reasoning	*n* (%) Not Promoting	Predominant Reasoning
0-2	3 (8%)	Judgmental	33 (92%)	Rule-based
3	9 (90%)	Both	1 (10%)	Judgmental
4-6	36 (95%)	Both	2 (5%)	Judgmental

Table 4d. Lenient assessors condition: Number and percentage of novice raters choosing to promote or not promote residents at various levels of performance, along with the predominant forms of reasoning used to justify their promotion decisions

# Successful EPAs	*n* (%) Promoting	Predominant Reasoning	*n* (%) Not Promoting	Predominant Reasoning
0-2	0 (0%)	--	38 (100%)	Rule-based
3	6 (46%)	X	7 (54%)	Judgmental
4-6	33 (87%)	Rule-based	5 (13%)	Judgmental

X = cannot be classified

Table 4e. Difficult personal circumstance condition: Number and percentage of novice raters choosing to promote or not promote residents at various levels of performance, along with the predominant forms of reasoning used to justify their promotion decisions

# Successful EPAs	*n* (%) Promoting	Predominant Reasoning	*n* (%) Not Promoting	Predominant Reasoning
0-2	3 (11%)	Judgmental	33 (89%)	Both
3	8 (67%)	Both	4 (33%)	Judgmental
4-6	28 (78%)	Both	8 (22%)	Judgmental

Table 4f. Unprofessional behaviour condition: Number and percentage of novice raters choosing to promote or not promote residents at various levels of performance, along with the predominant forms of reasoning used to justify their promotion decisions

# Successful EPAs	*n* (%) Promoting	Predominant Reasoning	*n* (%) Not Promoting	Predominant Reasoning
0-2	0 (0%)	--	36 (100%)	Both
3	4 (33%)	Rule-based	8 (67%)	Judgmental
4-6	21 (58%)	Rule-based	15 (42%)	Judgmental

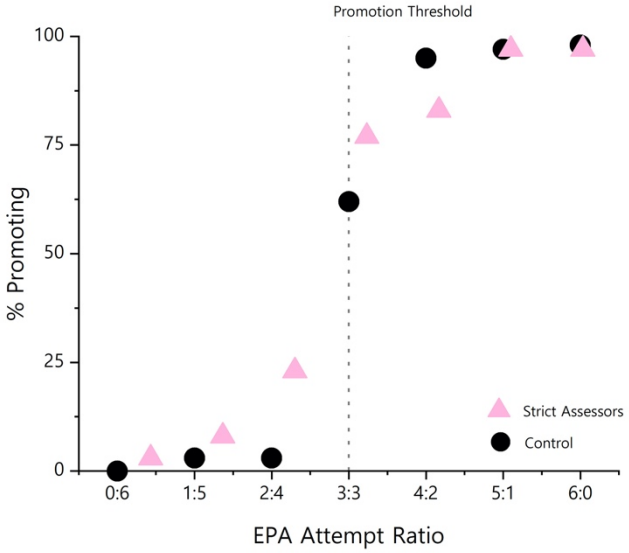

Figure 1a. Novice raters: Promotion decision curves for the experimental condition containing anecdotal evidence about strict assessors, and the control condition, which contained no anecdotal evidence

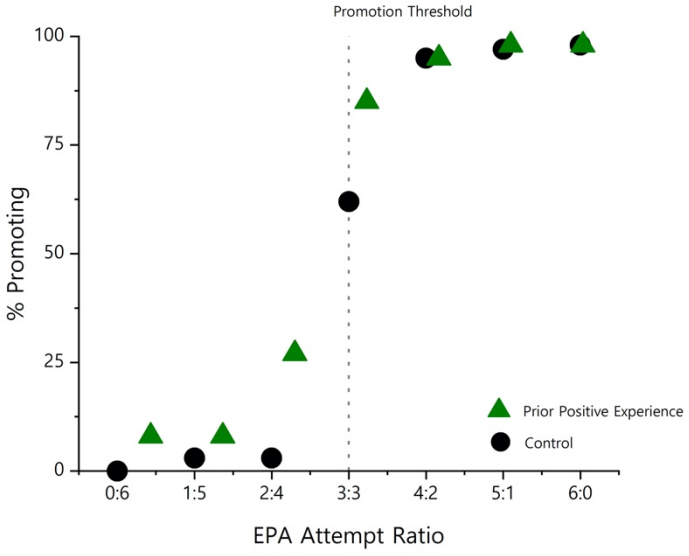

Figure 1b. Novice raters: Promotion decision curves for the experimental condition containing anecdotal evidence about a prior positive experience working with a resident, and the control condition, which contained no anecdotal evidence

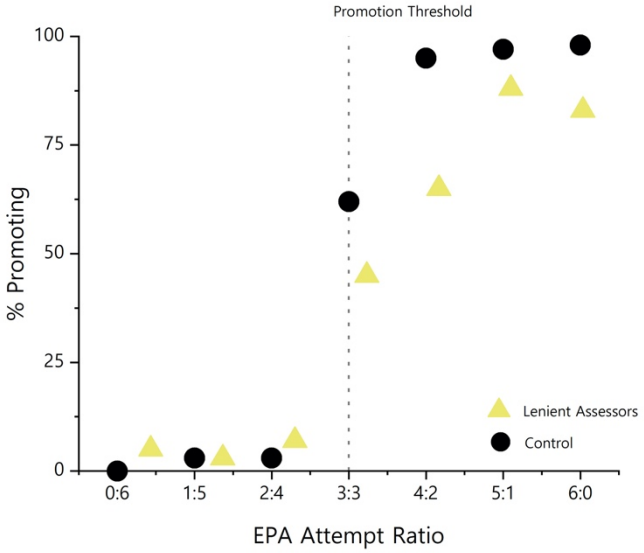

Figure 1c. Novice raters: Promotion decision curves for the experimental condition containing anecdotal evidence about lenient assessors, and the control condition, which contained no anecdotal evidence

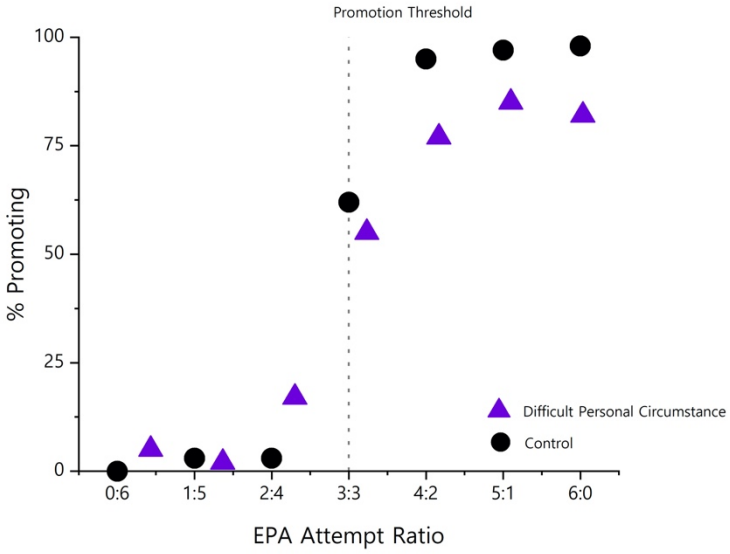

Figure 1d. Novice raters: Promotion decision curves for the experimental condition containing anecdotal evidence about a resident's difficult personal circumstances, and the control condition, which contained no anecdotal evidence

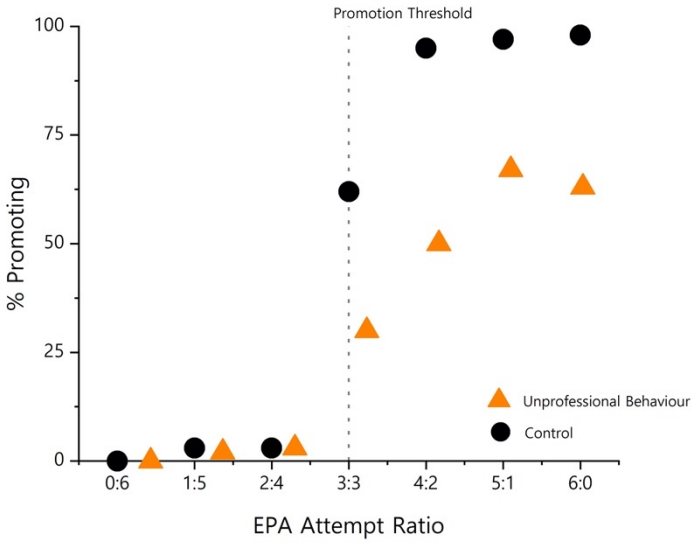

Figure 1e. Novice raters: Promotion decision curves for the experimental condition containing anecdotal evidence about a resident's unprofessional behaviour, and the control condition, which contained no anecdotal evidence

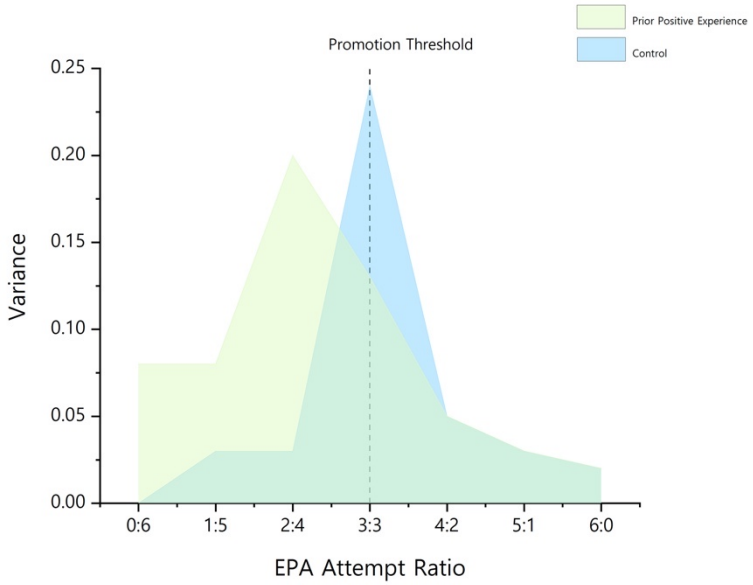

Figure 2a. Novice raters: Variance in promotion decisions when presented with a representative positive anecdote (i.e., prior positive experience), shown in green. The variance of the control condition, shown in blue, is presented for comparison

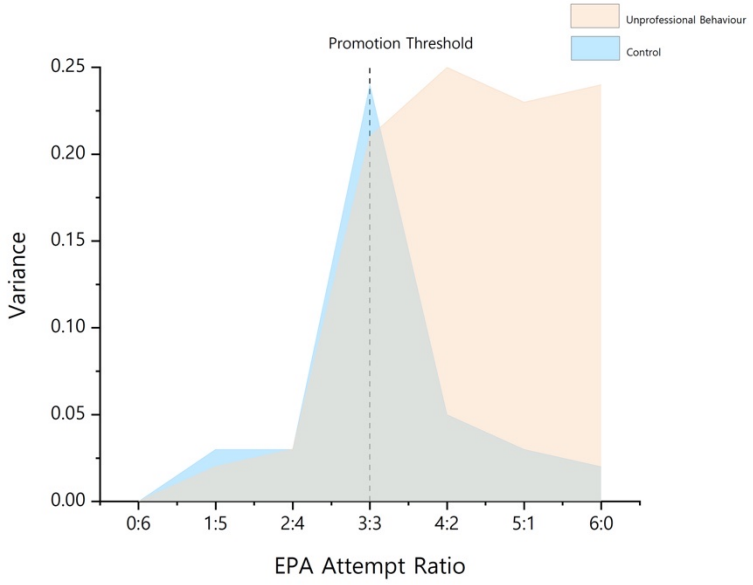

Figure 2b. Novice raters: Variance in promotion decisions when presented with a representative negative anecdote (i.e., unprofessional behaviour), shown in orange. The variance of the control condition, shown in blue, is presented for comparison

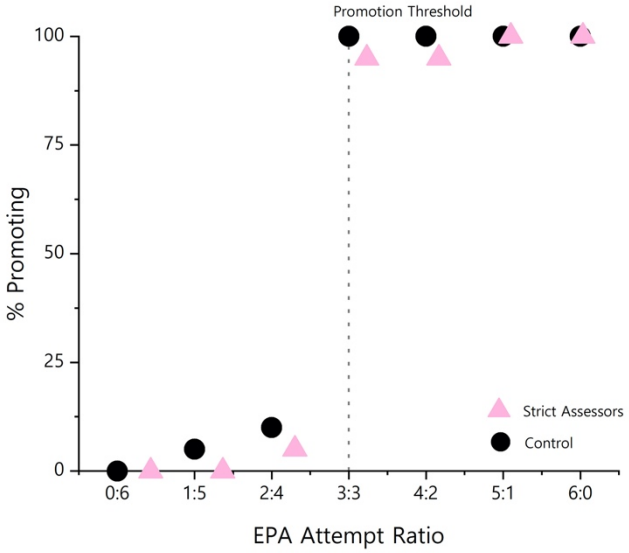

Figure 3a. Experienced raters: Promotion decision curves for the experimental condition containing anecdotal evidence about strict assessors, and the control condition, which contained no anecdotal evidence

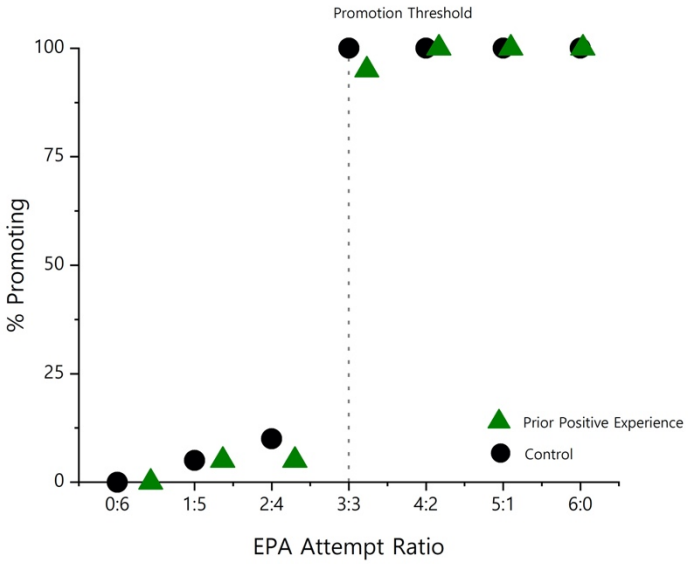

Figure 3b. Experienced raters: Promotion decision curves for the experimental condition containing anecdotal evidence about a prior positive experience working with a resident, and the control condition, which contained no anecdotal evidence

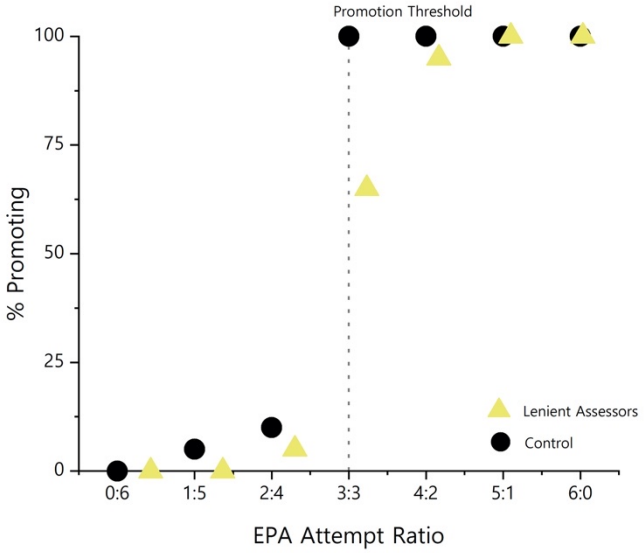

Figure 3c. Experienced raters: Promotion decision curves for the experimental condition containing anecdotal evidence about lenient assessors, and the control condition, which contained no anecdotal evidence

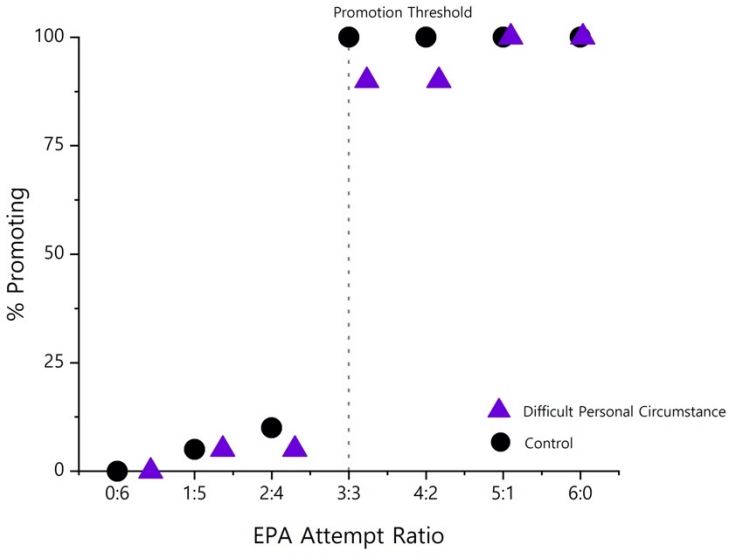

Figure 3d. Experienced raters: Promotion decision curves for the experimental condition containing anecdotal evidence about a resident's difficult personal circumstances, and the control condition, which contained no anecdotal evidence

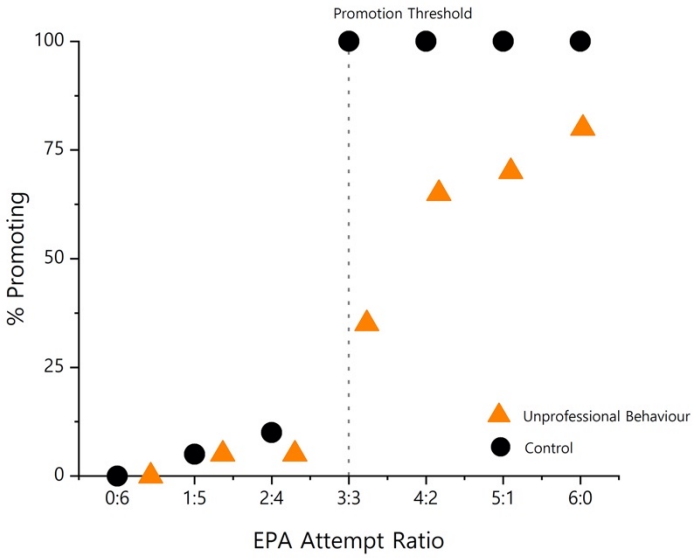

Figure 3e. Experienced raters: Promotion decision curves for the experimental condition containing anecdotal evidence about a resident's unprofessional behaviour, and the control condition, which contained no anecdotal evidence

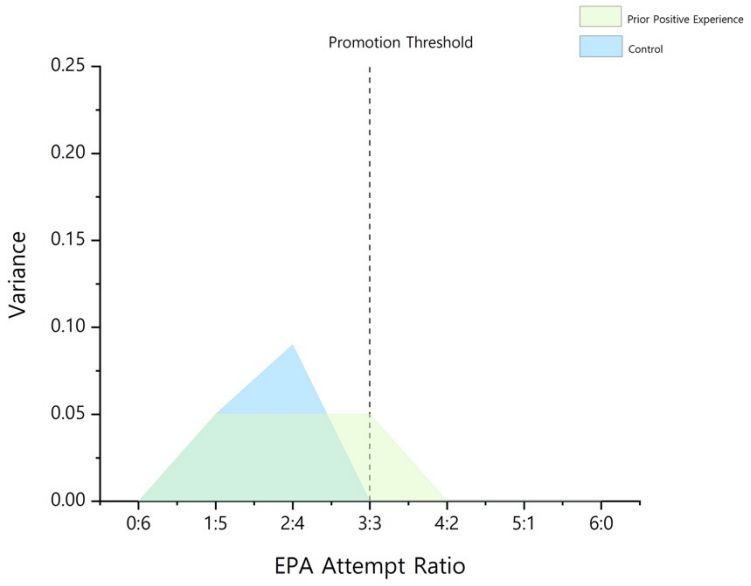

Figure 4a. Experienced raters: Variance in promotion decisions when presented with a representative positive anecdote (i.e., prior positive experience), shown in green. The variance of the control condition, shown in blue, is presented for comparison

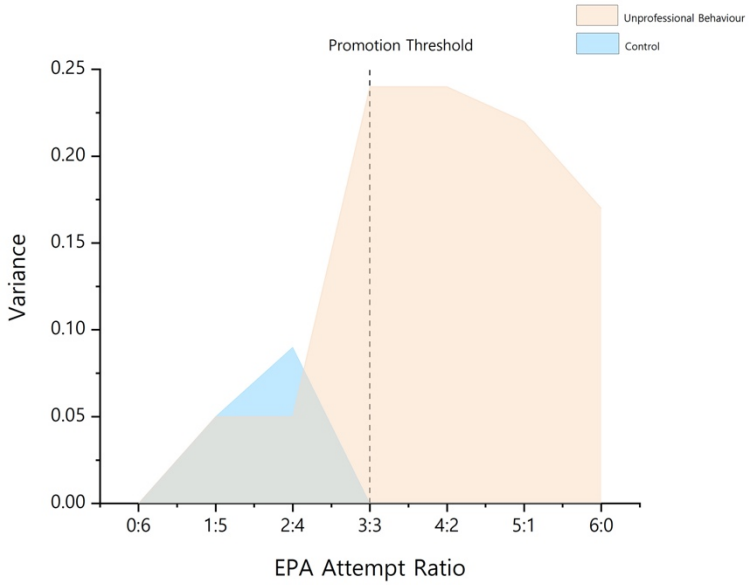

Figure 4b. Experienced raters: Variance in promotion decisions when presented with a representative negative anecdote (i.e., unprofessional behaviour), shown in orange. The variance of the control condition, shown in blue, is presented for comparison

Chapter 4

Understanding How Groups Make Promotion Decisions

4.1 Introduction

Although research on the decision-making processes of CCs is still in its infancy, research on group decision making has a rich history in psychology and organizational behaviour. One of the most widely-used theoretical frameworks is Davis' social decision scheme (SDS) theory, which conceptualizes group decision making as a combinatorial process by which individual member preferences are transformed into a group, or collective, response (Davis, 1973; Kameda, Tindale, & Davis, 2003; Kerr & Tindale, 2004). This process begins with individual member preferences, which depend on the unique values and experiences that group members bring to the table. Group composition refers to the distribution of individual member preferences within a group, such as the number of members who prefer one decision alternative over another. Through information sharing and discussion, individual preferences are combined into a group response using rules known as decision schemes. Examples of decision schemes include majority/plurality, proportionality, and equiprobability, with majority/plurality schemes being by far the most common (Kerr, MacCoun, & Kramer, 1996; Stasser, 1999).

SDS theory been applied to groups in a variety of contexts, both to explain existing experimental findings and to make predictions based on theoretical assumptions (Stasser, 1999). One of the most ubiquitous findings from studies using this framework is that in most situations, especially those in which there is not an 'optimal' or 'correct' answer, groups will choose the decision alternative that is supported by the majority of their

members (Kameda et al., 2003; Kerr & Tindale, 2004). The power of a strength-in-numbers approach to decision making is apparent among both humans and animals. For example, honeybees will form swarms around sites with optimal nesting conditions and perform waggle dances as a way of advertising them to others (Seeley, 2010). Other bees are highly responsive to these cues and will preferentially select sites with large swarms as opposed to those with only one or two other visitors (Seeley & Buhrman, 1999). Similar behaviours can be observed among humans. For example, it is much more likely that someone will purchase a new item if everyone else is doing it too, as demand for a product is a potential signal of its quality and/or value (Ramya & Ali, 2016).

While this approach to decision making is certainly not foolproof as it can lead to a self-fulfilling prophecy, studies have shown that, on average, majority/plurality schemes provide the best compromise between decision-making accuracy and efficiency. In a study comparing nine different decision schemes, Hastie and Kameda (2005) found that majority/plurality schemes could outperform a group's best individual member and were roughly equal in accuracy to more complex (and therefore less efficient) models, such as those involving weighted averaging based on past performance. This may help explain why majority/plurality schemes are so prevalent among both human and animal groups (Hastie & Kameda, 2005; Kameda, Wisdom, Toyokawa, & Inukai, 2012).

An important implication of majority/plurality decision schemes is that the distribution of individual member preferences within a group will have a large impact on the decisions that it makes (Stasser, 1999). If an unfavourable cognitive bias is highly prevalent among group members (e.g., >50% in a simple majority scheme), then the

group would be expected to amplify the bias, leading to a worse outcome than if the decision had been made individually (Kerr et al., 1996). If, however, the opposite is true and the cognitive bias is prevalent among only a minority of members, then the group would be expected to confer an advantage (Kerr et al., 1996). Thus, the same group processes can lead to both good and bad decisions, depending on the distribution of initial preferences within the group (Tindale, Smith, Dykema-Engblade, & Kluwe, 2012).

SDS theory and its emphasis on majority/plurality decision schemes also help to explain group polarization, the tendency of groups to make decisions that are more 'extreme' than the preferences of their individual members (Isenberg, 1986; Myers & Lamm, 1976). Initially, scientists struggled to understand why some groups made riskier decisions than individuals while others were more cautious (Myers & Lamm, 1976). A common hypothesis was that some group members may be more persuasive than others; however, subsequent research using SDS theory showed that it was not persuasive argumentation that mattered, but the number of members that preferred a particular decision alternative (Kameda et al., 2003; Kerr, Davis, Meek, & Rissman, 1975; Zuber, Crott, & Wener, 1992). If the majority of members preferred the risky alternative, then the group would become polarized in that direction; conversely, if the majority preferred the cautious alternative, then the opposite would occur.

4.1.1 Information Processing Model of Groups

While SDS theory is helpful for describing phenomena commonly observed among groups, it provides little insight into the mechanisms that underpin group decision-making processes. For insight into some of these mechanisms, one must turn to other

models, such as the information processing model of groups (Hinsz, Tindale, & Vollrath, 1997). This model suggests that groups can outperform individuals because groups have the capacity to share more information. In addition, since more people are involved in the decision-making process, groups afford better opportunities to scrutinize this information (i.e., by virtue of having more people involved in the decision-making process; Hinsz et al., 1997; Mesmer-Magnus & DeChurch, 2009; Stasser, 1999).

The idea of groups as information processors became popular in the late 1990s, although it had been circulating in the literature for years prior. For example, several early scholars had alluded to or explicitly recognized groups as information processors (Driver & Streufert, 1969; Miller, 1960; Von Cranach, Ochsenbein, & Valach, 1986). However, the idea was perhaps most famously reinforced through Stasser and Titus' (1985) hidden profile paradigm. Using an experimental task in which some but not all information was shared among group members, the authors found that information common to multiple group members was more likely to be discussed than information held by only one member, thereby leading to poorer decisions. This became known as the common knowledge effect and has spurred hundreds of follow-up studies that have underscored the importance of robust information-sharing practices in ensuring successful group outcomes (Mesmer-Magnus & DeChurch, 2009; Tindale & Sheffey, 2002).

Hinsz and colleagues (1997) defined information processing in groups as, "the degree to which information, ideas, or cognitive processes are shared, and are being shared, among the group members and how this sharing of information affects both individual- and group-level outcomes" (p. 43). They proposed a model of group decision

making that drew upon concepts central to cognitive psychology such as attention, encoding, storage, retrieval, and processing to explain how groups make decisions. While it was understood that groups did not necessarily operate in exactly the same way as individuals, these concepts were helpful metaphors for understanding group decision making and helped increase the popularity of the model (Hinsz et al., 1997).

The information processing model of groups consists of multiple components. Every group is embedded in a particular context, which influences all aspects of their functioning, including processing objectives, attention, encoding, and so on (Hinsz et al., 1997). Processing objectives refer to a group's overall goals, which can be based on "instructions, task characteristics, procedural factors, member perspectives, roles, and norms" (Hinsz et al., 1997, p. 45). These objectives will lead the group to attend to specific types of information and not others. Incoming information moves in and out of memory via storage and retrieval processes. The processing workspace allows incoming information from attentional and/or retrieval processes to be integrated and processed on the basis of different rules, strategies, and procedures. This results in a group response, which can take the form of a decision, inference, opinion, or solution. Group responses often lead to feedback, which can inform the group's future actions (Hinsz et al., 1997).

The information processing model of groups has also been used to unify previous literature on group decision making. Hinsz and colleagues (1997) proposed a four-factor model, with each factor linked to a key aspect of the group literature. For example, commonality-uniqueness of information links to prior research on information such as Stasser and Titus' (1985) work on hidden profiles; convergence-diversity of ideas links to

research on shared mental models, which are a prevalent concept in the group decision-making literature (Hauer et al., 2016; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000; Mohammed & Dumville, 2001); accentuation-attenuation of cognitive processes links to SDS theory; and belongingness-distinctiveness of members links to the role of social processes in group decision making, which will be further explored in Chapter 5. Tindale and Kameda (2000) attempted to build on this model by proposing the concept of social sharedness, which they suggested could be a unifying theme for understanding groups as information processors. The extent to which individuals share information—be it attitudes, identities, or cognitions—greatly affects group outcomes and underscores the importance of Hinsz and colleagues' (1997) model.

4.1.2 Study Purpose and Overview

The purpose of this study was to examine how groups make promotion decisions and whether this process confers any advantages over individual decision making. As in Chapter 3, this was examined in the context of CC decision-making processes and anecdotal evidence. Experiment 1 used a sample of novice raters to compare decisions made in groups with those made individually using the same simulated resident files as in Chapter 3. Experiment 2 explored the same question using a sample of experienced raters. Ongoing observations revealed that in practice, anecdotal evidence appeared to have a broader role than was captured by the experimental findings. The final portion of this study analyzed CC meeting and interview transcripts to explore the additional functions that anecdotal evidence may have in naturalistic decision-making environments.

4.2 Methods

4.2.1 Experiment 1: Novice Raters

Participants in this study were undergraduate students enrolled in the psychology participant pool at McMaster University in Hamilton, Ontario, Canada. During recruitment, participants were screened to ensure that they did not have any experience making promotion decisions prior to taking part in the study. Recruitment was done through SONA, an online system that allows students to complete research studies in exchange for course credit. Participants received two course credits in exchange for completing this study. Gigone and Hastie (1997) suggested that researchers should aim for equivalence between the number of participants that complete a study individually and the number that complete it in groups. Since 60 novice raters participated in the individual decision-making experiment described in Chapter 3, the present study aimed to recruit 12 groups of five participants.

4.2.1.1 Materials

The same 42 simulated resident files were used as in Chapter 3.

4.2.1.2 Procedures

This experiment consisted of two parts. In Part 1, novice raters completed the study individually using the procedures outlined in Chapter 3 Experiment 1. In Part 2, the same raters were invited into the lab and completed the task again, this time in groups of approximately five. When participants entered the lab, they were greeted by the primary investigator and asked to introduce themselves to their fellow group members. Next, they were read a list of study instructions that outlined the task and covered ground rules for

working as a group (e.g., mobile phones should be switched off or silent, disagreements should be voiced respectfully, only one participant should speak at a time). Participants were then given up to an hour to review the 42 cases and decide whether or not each resident should be promoted to the next stage of training. They were encouraged to come to a decision by consensus; however, the group could decide to take a vote if they were at an impasse. During the group decision-making task, the primary investigator and an undergraduate research assistant audio-recorded the conversation and took notes on what was happening in the room as the experiment unfolded.

4.2.1.3 Data Analysis

The permutation testing procedure described in Chapter 3 was used to compare decisions made by groups with decisions made by individuals. Only overall comparisons were made because an effect of groups was expected at all performance levels, not just at the midpoint of each curve. In fact, an effect would have been least probable at the midpoint, since the findings of Chapter 3 showed that this is where the effect of anecdotal evidence was most salient. Following the permutation tests, the variance of decisions made in groups and the variance of decisions made individually were calculated for each condition, at each level of performance, and compared using paired *t*-tests.

4.2.2 Experiment 2: Experienced Raters

Participants in this study were CC members with approximately two years of involvement in resident promotion and review processes. Chairs of existing CCs at McMaster University in Hamilton, Ontario, Canada were sent a letter of information about the study to discuss with their members. If a committee wished to participate, a date

and time was arranged for the primary investigator to conduct the experiment; this usually occurred at the start or end of a regularly scheduled CC meeting. Since 20 experienced raters participated in the individual decision-making experiment described in Chapter 3, the present study aimed to recruit four CCs with approximately five members each.

4.2.2.1 Materials

The same 42 simulated resident files were used as in Chapter 3.

4.2.2.2 Procedures

The same procedures were used as in Experiment 1, except participants did not complete the experiment online before making group decisions because many would have previously completed Chapter 3 Experiment 2. Thus, decisions made by CCs (i.e., in groups) were compared to those made by individual raters in Chapter 3 Experiment 2.

4.2.2.3 Data Analysis

The same data analysis procedures were used as in Experiment 1.

4.2.3 Anecdotal Evidence in Naturalistic Decision-Making Environments

Data collected during the implementation study described in Chapter 2 were used in this study, which constituted a separate analysis. Of the 9 CCs observed, four consented to have their meetings audio recorded and transcribed; their characteristics are presented in Figure 1, along with the amount of available data for each. In addition, the interviews conducted during Phase 3 of the implementation study included questions to gauge CC members' perceptions of and experience with anecdotal evidence. The perspective of individual members was relevant not only because individual members' preferences constitute important part of the group decision-making process (Stasser,

1999), but also because the questions asked members to reflect on the role of anecdotal evidence in the group decision-making processes in which they had taken part. Triangulation of evidence from both data sets allowed for a better understanding of how CCs use anecdotal evidence in naturalistic decision-making environments.

4.2.3.1 Data Analysis

Analysis of the meeting and interview transcripts took place using Braun and Clarke's (2006) six-step approach to thematic analysis: familiarizing oneself with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing a written account of the data. As a PhD candidate in psychology and health professions education, the primary investigator brought an external perspective to this work as she was not a member of any of the CCs that participated in this research. Strategies to maintain the trustworthiness of the qualitative findings included prolonged engagement with the data, triangulation between qualitative and quantitative data, reflexive journaling, clear documentation of the analytic process, and peer debriefing (Nowell, Norris, White, & Moules, 2017).

4.2.4 Ethics Approval

The studies outlined in this chapter were approved by the Hamilton Integrated Research Ethics Board (HiREB-4534).

4.3 Results

4.3.1 Experiment 1: Novice Raters

Sixty-six undergraduate students (48 females, 18 males) participated in this experiment. The majority of participants were in first year (n = 42; 64%), and their most

common programs of study were life sciences (n = 21; 32%) and health sciences (n = 8; 12%). Participants did not report any prior experience with making promotion decisions prior to participating in the study, confirming that the sample consisted of novice raters.

Of the 66 participants, all but two completed the individual phase of the experiment (i.e., Part 1). Both of these participants still attended the group phase of the experiment (i.e., Part 2). Additionally, three participants who completed Part 1 of the experiment did not attend Part 2. Thus, there were 63 participants in Part 2 of the experiment, for a total of 13 groups (the intended 12 groups plus an additional group as a backup, in case any unexpected issues arose). All groups had five participants except for two, which had three and four participants due to the forementioned no-shows.

Figures 2a to 2f compare novice raters' individual promotion decisions to those made by groups for each condition. Groups were significantly less sensitive to anecdotal evidence than individuals in all conditions except the unprofessional behaviour condition and the 3:3 subcondition of the lenient assessors condition, in which they were *more* sensitive to anecdotal evidence than individuals (Table 1). Despite being less likely than individuals to use anecdotal evidence in all but these conditions, some groups could still be observed using this information when faced with high levels of ambiguity, such as in the 3:3 subcondition (i.e., three successful: three unsuccessful EPA attempts).

Figures 3a to 3f compare the variance of novice raters' individual promotion decisions to that of groups for each condition. Decisions made by groups were significantly less variable than decisions made by individuals in the prior positive experience, lenient assessors, and difficult personal circumstance conditions (Figures 3c

to 3e; Table 2). The control and strict assessors conditions displayed a similar pattern; however, differences in variance were not statistically significant (Figures 3a and 3b; Table 2). The unprofessional behaviour condition was once again an anomaly, as both individual and group decisions were highly variable in this condition (Figure 3f).

4.3.2 Experiment 2: Experienced Raters

Five CCs from across five specialties participated in this experiment (once again, an extra committee was included as a backup, in case any unexpected issues arose). Committees ranged in size from three to nine members, with the average number of members being five. All but two members were clinical faculty, with an average of 12 years in practice, 1.5 years of experience on a CC, and 4.5 years of experience on a promotion committee other than a CC, such as a residency program committee.

Figures 4a to 4f compare experienced raters' individual promotion decisions to those made by groups for each condition. Experienced raters were less sensitive to anecdotal evidence than novices, and this effect was further attenuated when experienced raters made decisions in groups as opposed to individually (Table 3). Nonetheless, some conditions still elicited a reliance on anecdotal evidence—most notably the unprofessional behaviour condition (Figure 4f). It is noteworthy that in this condition, the effect was in the opposite direction to novices: groups were *less* sensitive to anecdotes about unprofessionalism than were individuals.

Figures 5a to 5f compare the variance of experienced raters' individual promotion decisions to that of groups for each condition. Although differences in variance between groups and individuals were not statistically significant, the graphs show a pattern of

reduced variability in all conditions except unprofessional behaviour (Table 4). Aside from at the midpoint (i.e., the 3:3 subcondition), the variability in each condition was often reduced to zero when promotion decisions were made in groups.

4.3.3 The Role of Anecdotal Evidence in Naturalistic Decision-Making Environments

Audio recordings of CC meetings totaled 17 hours across four programs, with a detailed breakdown by committee available in Figure 1. Of the 20 individuals contacted, 12 (60%) responded to an invitation to participate in the interview. These participants represented CCs across seven specialties, and a range of membership types (clinical faculty: $n = 7$, PhD-trained researchers: $n = 1$, residents: $n = 4$). Seven participants reported serving on one CC while three reported serving on two or more CCs. Participants had attended an average of six CC meetings at the time of the interview.

4.3.3.1 Prevalence of Anecdotal Evidence

Interviews with CC members suggested that anecdotal evidence was highly prevalent in CC decision-making processes. One member stated: “It actually happens a fair bit” (PhD-Trained Researcher). Another remarked, “It’s always there. … Inevitably, whenever people meet in this kind of setting, if they have any sort of anecdotal experience to share, they will” (Resident). One member even went as far as to say that, as a resident, *all* of their contributions were anecdotal because: “We don’t have much other evidence. We don’t have access to the evaluations. … The residents are there to support the other residents and to provide any context that they might be able to” (Resident). In total, nine out of the 12 participants stated that they had encountered anecdotal evidence while serving as a CC member and could recall specific examples.

Analysis of observation notes and meeting transcripts corroborated these findings. It was possible to identify examples of each type of anecdotal evidence featured in the vignettes presented in Experiments 1 and 2. Table 5 provides real-world examples of anecdotal evidence corresponding to each vignette.

4.3.3.2 Functions of Anecdotal Evidence

Five different functions of anecdotal evidence were identified and classified by the primary investigator: substituting, corroborating, explaining, questioning, and prompting additional action. These functions were not mutually exclusive, as anecdotal evidence could often serve more than one purpose.

Substituting was a direct function of anecdotal evidence in that it could sometimes be used to make decisions in lieu of missing assessment data. This rarely occurred at the level of stage promotions; however, it was common for CCs to use anecdotal evidence to deem a resident competent on skills for which they were missing one or two of the required observations. One member noted:

> "There's a lot of missing information. There were certainly times when … the residents were ready to move to the next level, but the documentation wasn't there. And so, the anecdotal information was, 'Okay, we know that they've already completed it and, you know, they just haven't had the information entered into the database" (PhD-Trained Researcher)

The observations and interviews also revealed a number of indirect, or 'hidden,' functions of anecdotal evidence. Corroborating occurred when anecdotal evidence was used to support the decision indicated by existing assessment data. For example, if a resident's file pointed to consistently above-average performance, members still sometimes shared their own positive experiences working with that

resident. The opposite could also occur for residents whose performance was consistently below average. In both cases, anecdotal evidence appeared to serve as form of validation to signal that the committee's decision making was appropriate. According to one participant: "The anecdotes either confirm what the sense already is around the table or what the right decision should be" (Faculty).

Explaining occurred when residents exhibited patterns of performance or behaviour that were difficult to interpret. During one CC meeting, members were concerned about a resident who appeared to be strong clinically but lacked initiative. Members were conflicted about the situation since on one hand, it felt unfair to hold a resident with strong clinical skills back, but on the other, they were concerned about their ability to be successful in the long run. The committee's decision making was made easier when the program director informed committee members that they had some information about the resident's health status that could help explain their perceived lack of initiative. After learning this new information, the committee felt comfortable promoting the resident on the condition that appropriate supports were in place to ensure their success.

It is noteworthy that in the majority of cases in which anecdotal evidence served an explanatory function, it was usually a single person, such as the program director, providing the information. The full set of information in a resident's file was not always accessible to all CC members, especially if the committee was new and did not yet have fully operational data sharing mechanisms in place. During one meeting, this created an interesting tension when a committee member questioned

the defensibility of a decision because the committee had relied upon a secondhand summary of critical information from the program director:

> *Member:* "If this is going to be the decision, though, I worry that we don't have enough documentation to make it defensible at this point in time."
>
> *Program Director:* "Um … well, I think we do … when we add this to what's in their file … and trust me, there's … at least 30 pages of letters … I mean, the documentation is there in that file if this is the decision…"
>
> *Member:* "Yeah, but … the committee hasn't seen that file…"

Questioning tended to occur in the context of anecdotal evidence that prompted committee members to scrutinize a piece of assessment data that they would otherwise have taken at face value: "If there's a major disagreement between the assessment data and the anecdotal evidence, then it might bring into question some of the … data and how it was gathered" (Faculty). This often occurred when a member shared insider information about a resident or assessor with the rest of the committee. During one meeting, a committee was reviewing a resident's performance on a clinical examination when the chair revealed that they were present at the exam and noticed that one of the assessors was being unreasonably harsh in their evaluations. This caused the committee to question the credibility of data from this assessor and altered their decision making by putting less emphasis on these data when making promotion decisions.

A final function of anecdotal evidence in CC decision making was prompting additional action. This often occurred in conjunction with the other functions noted above, such as explaining or questioning, and typically took the form of seeking more evidence, following up with appropriate parties, and/or

creating a plan for supporting the resident going forward: “I do think it might influence our encouragement to that person completing their EPAs. So, I think when we go back and we’re working with them, we’re sort of like, ‘Let me do some of those for you,’ or ‘Let me get you ahead on that’” (Faculty).

4.3.3.3 Perceptions of Anecdotal Evidence

This study also investigated CC members’ perceptions of anecdotal evidence, and in particular, how it should (or should not) be used when making group-based promotion decisions. Although the focus was on individual members, their views are likely to reflect the conversations that would occur at the group level should a CC encounter anecdotal evidence. An interesting finding was that despite its prevalence, most CC members viewed anecdotal evidence as lacking in credibility and something to be avoided whenever possible. One member stated: “The anecdotal stuff is going to be the most problematic … so, therefore, you should try to get away from it as much as possible” (Faculty). Another stated: “I think we have to be very careful not to give [it] too much weight … you know, there’s that saying, right … the plural of anecdotes is not data” (Faculty).

A specific concern about anecdotal evidence had to do with its fairness to residents: “I would be worried that this kind of information is not fair for everyone, right? Because then it would maybe depend on who was on the committee and who’s been in touch with which resident” (Faculty). Along similar lines, another member warned: “Just be careful with anecdotal evidence because negative anecdotes will always overpower positive anecdotes” (Resident).

Nonetheless, a few participants saw value in anecdotal evidence. One resident stated, "I'm a big fan of anecdotal information, actually, which maybe other people aren't—I don't know" (Resident). Another member shared:

> "I actually think it's useful. … A resident may not be sure how to enter [assessment data] or may have entered in the wrong information or may have personal issues … And without that anecdotal information you don't know why and you're going to hold somebody back when it's really not necessary or you might be pushing somebody through even though they're not quite ready" (PhD-Trained Researcher)

4.3.3.4 Strategies for Handling Anecdotal Evidence

The observations did not uncover any formalized strategies for dealing with anecdotal evidence, although some members did encourage the committee to ensure that it was appropriately documenting its decisions and any evidence used to come to those decisions. Non-clinical members, such as PhD-trained researchers, could play an important role in this process, as they saw keeping the committee accountable for its actions as part of their role:

> "I'm just trying to think of the legality of everything. You never want to hold somebody back and not have it documented, so all pieces of information should be documented wherever possible so you can provide good rationale for what the decisions were" (PhD-Trained Researcher)

Members also discussed other ways in which anecdotal evidence could be handled at the committee level, including looking for patterns of evidence rather than singular data points. As one member put it, "If you have multiple people and multiple anecdotes, that's no longer anecdotes, those are now habits" (Resident). Other members suggested that while anecdotal evidence could sometimes be useful for providing context, it should be given little to no weight when making promotion

decisions: “We actually don’t include it in the discussions for promotion … and I don’t think you should” (Faculty). Finally, one member pointed out the importance of thoroughly scrutinizing any anecdotal evidence that is received: “I think it should be carefully included. You have to look at who the anecdotal evidence is about, but also who is providing the evidence” (Resident).

4.4 Discussion

This study investigated the role of anecdotal evidence in novice and experienced raters’ group-based promotion decisions and compared decisions made by groups to decisions made by individuals. Experiment 1, in which novice raters made promotion decisions about simulated files both individually and in groups, showed that groups were less sensitive to anecdotal evidence than individuals in all conditions but one (unprofessional behaviour). They also significantly reduced the variability among individual raters’ decisions. Experiment 2 used the same paradigm with experienced raters and found similar results; groups were once again less sensitive to anecdotal evidence than individuals and variability was reduced among groups, although this finding was not statistically significant. A separate analysis of CC meeting transcripts and interviews revealed not only a direct role of anecdotal evidence in substituting for missing assessment data, but also a number of indirect, or hidden, functions, including corroborating, explaining, questioning, and promoting additional action.

4.4.1 The Role of Groups in Making Promotion Decisions

The findings of this study showed that groups significantly reduced the variability among experienced and novice raters’ individual promotion decisions; however, they

were still responsive to anecdotal evidence in certain situations, such as when cases were highly ambiguous, or anecdotes related to unprofessional behaviour. This suggests that groups may help reduce unwanted variability among individual raters while still remaining sensitive to variability that is meaningful and may signal something important, such as data that are 'problematic' (Pack, Lingard, Watling, Chahine, & Cristancho, 2019) or future performance concerns (Odorizzi et al., 2020; Schumacher et al., 2018). This is important because it helps support the use of CCs and other promotion committees in medical education. Not only do groups make more consistent promotion decisions, which helps to increase their defensibility, having more individuals involved may also increase stakeholder buy-in and engagement in CBD. A good example comes from Chapter 2, where residents involved as CC members had much greater confidence in their programs' promotion and review processes than those who were not.

The group decision-making literature suggests that groups have different ways of reconciling variability among individual raters. A crude mechanism may be a 'strength in numbers' approach (Kameda et al., 2003; Kerr & Tindale, 2004; Stasser, 1999). The fact that multiple members favour a particular decision alternative is a sign that it may be the best option, which is why groups of both humans and animals often make decisions using majority/plurality decision schemes (Gigone & Hastie, 1997; Kameda et al., 2003; Tindale & Kameda, 2000). This decision scheme likely also applies to this study, as the effect of anecdotal evidence was accentuated in groups when it was preferred by the majority of members (i.e., unprofessional behaviour) but attenuated when it was preferred by only a minority of members (all other conditions).

However, a problem with majority/plurality decision schemes is that when used uncritically and without thorough discussion, they can sometimes lead to poor decision-making outcomes. Janis' (1982) concept of groupthink suggests that cohesiveness, or the desire for cohesiveness, can lead group members to become overly agreeable in an attempt to minimize conflict. The consequence of this approach is a lack of critical evaluation, which can result in poor decision-making outcomes. One wonders if a similar phenomenon might have occurred in this study when experienced raters became less sensitive to anecdotal evidence about unprofessionalism when working in groups as opposed to individually. As suggested by the interview findings and corroborated by the literature (e.g., Pack et al., 2019), CC members are often uncomfortable with the idea of using 'unobjective' data to make decisions. Working as a group could have exacerbated this tendency and resulted in the promotion of residents with professionalism concerns. While in the real world CCs would undoubtedly have more tools at their disposal to handle such a situation (e.g., they could ask their colleague to document the situation or search for additional evidence), the idea that excessive conformity can lead to poor decisions remains an important pitfall of which CCs need to be aware, especially given the tendency for groups to rely on a 'strength in numbers' approach.

Aside from highlighting the importance of sharing and scrutinizing information as opposed to simply following a set of predetermined decision rules, this study also highlights the importance of ensuring diverse representation on CCs and other decision-making bodies (Page, 2007). Given the commonality of a 'strength in numbers' approach, it is important that multiple and varied perspectives are reflected on the committee

(Bowers, Pharmer, & Salas, 2000; Stasser, 1999). Having too many committee members with similar backgrounds and experiences could exacerbate the previously mentioned issues of conformity because information would be shared and scrutinized using only a singular cognitive lens (Jackson, May, & Whitney, 1995). It could also lead to a situation in which minorities uncomfortable speaking up because they do not feel as if their opinions matter against a strong majority (Elsass & Graves, 1997). The extent to which different members feel included in the decision-making process and how this influences their interactions is a critical determinant of the strength of information sharing within CCs and will be explored in Chapter 5.

4.4.2 The Role of Anecdotal Evidence in Information Processing

In addition to revealing important characteristics of how CCs make decisions, this study also provided a more in-depth understanding of how anecdotal evidence is processed at the group level. The experimental findings showed that anecdotal evidence can act as a source of information about a resident's performance, particularly when other performance data are difficult to interpret. This was also observed in practice when CCs used anecdotal evidence to substitute for missing assessment data. However, the observations and interviews conducted as part of this study also revealed a number of indirect, or 'hidden,' functions of anecdotal evidence, such as corroborating, explaining, questioning, and prompting additional action. Rather than acting as a source of information, anecdotal evidence was seen in these cases to modulate the information processing cycle as a whole.

One way in which anecdotal evidence could modulate CCs' information processing cycle is by acting as a 'flag' to guide the committee's attention (Hinsz et al., 1997). In larger committees, it was common for committee members to spend only a few minutes on each file; however, if a concerning anecdote was shared, CCs would be prompted to spend much more time on discussion. A common strategy was to look for evidence in the resident's file that could either confirm or deny the anecdote (e.g., by taking a closer look at the qualitative comments on their evaluations or having the program director share any formally documented concerns). Upon hearing an anecdote from one member, some chairs would also encourage other members to share their perspectives as a way of triangulating evidence across multiple members. Thus, in addition to redirecting the committee's attention, anecdotal evidence can also spur the retrieval of other information to help guide the committee' decision-making process (Hinsz et al., 1997).

From an information processing lens, anecdotal evidence also helped CCs encode information, which "involves the structure and interpretation of information in groups" (Hinsz et al., 1997, p. 47). As previously described by Pack and colleagues (2019), it is not uncommon for CCs to encounter 'problematic evidence' that is difficult to interpret. Examples from the observations in this study included situations in which assessment forms were not filled out correctly or when data did not match the committee's experiences with a resident, such as the example of a talented resident who was not meeting their potential. In these cases, anecdotal evidence could provide important contextual information that allowed committees

to more easily make sense of assessment data, such as learning that the resident's apparent lack of motivation could be explained by health-related challenges.

Anecdotal information could also influence other parts of the information processing cycle. Even if a committee chose not to act on a piece of anecdotal evidence immediately, ensuring that it was appropriately documented ensured that it is available in 'storage' should additional, similar information come to light in the future (Hinsz et al., 1997). Anecdotal evidence could also prompt groups to alter their responses and/or engage in a feedback loop (Hinsz et al., 1997). For example, committees could take additional action above and beyond making promotion decisions, such as gathering additional information or ensuring that the promotion decision is linked with a robust enhanced educational or remediation plan. They could also discuss ways to improve the quality of their decision-making processes in the future, such as by ensuring that certain information was available to all committee members. These additional functions of CCs are important to note, as they may not always be captured when thinking about these committees purely as summative decision-making bodies (Pack, Lingard, Watling, & Cristancho, 2020).

4.4.3 Limitations and Future Directions

Some of the limitations discussed in Chapter 3, including using undergraduate psychology students as novice raters and the need to simplify resident files in order to isolate anecdotal evidence as a variable, also apply to this study. Moreover, undergraduate students may not have been as familiar with participating in group decision-making processes as experienced raters. Another

limitation is that while some of the experienced raters who completed the individual decision-making experiment in Chapter 3 were members of the CCs who completed the experiment as a group, detailed information on the number of participants that overlapped between the two experiments is not available as the experiment in Chapter 3 was anonymous. Finally, only data from four CCs at a single academic centre were used to investigate the role of anecdotal evidence in naturalistic decision-making environments. While this is more than in most other studies on CCs, which are mostly program-specific, it may not represent the experience of all CCs, particularly since three of the four CCs represented surgical programs. Although no major differences were noted between these committees and the fourth committee, which represented a medical program, future studies may wish to compare these findings to those from other programs and sites.

4.5 Conclusion

The purpose of this study was to examine how groups make promotion decisions and whether this process confers any advantages over individual decision making. As in Chapter 3, this was examined in the context of CC decision-making processes and anecdotal evidence. Findings suggested that groups can help reduce unwanted variability among individual raters' promotion decisions while remaining sensitive to variability that may be meaningful; however, attention to information sharing and scrutiny and group composition is critical for ensuring that this benefit of groups is realized. A separate analysis of CC meeting transcripts and interviews revealed not only a direct role of anecdotal evidence in substituting for missing assessment data, but also a number of

indirect, or hidden, functions, including corroborating, explaining, questioning, and prompting additional action. These findings suggest that anecdotal evidence act as both an information source and as a moderator of how CCs process other information by guiding attention, facilitating the retrieval of other information, aiding in encoding, and prompting additional actions. Since social hierarchies and power can also play an important role in facilitating or inhibiting information processing in groups, the next chapter will explore the role of social processes in CC decision making.

4.6 References

Bowers, C. A., Pharmer, J. A., & Salas, E. (2000). When member homogeneity is needed in work teams: A meta-analysis. *Small Group Research, 31*(3), 305-327. https://doi.org/10.1177/104649640003100303

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Davis, J. H. (1973). Group decision and social interaction: A theory of social decision schemes. *Psychological Review, 80*(2), 97-125. https://doi.org/https://doi.org/10.1037/h0033951

Driver, M. J., & Streufert, S. (1969). Integrative complexity: An approach to individuals and groups as information-processing systems. *Administrative Science Quarterly, 14*(2), 272-285. https://doi.org/10.2307/2391105

Elsass, P. M., & Graves, L. M. (1997). Demographic diversity in decision-making groups: The experiences of women and people of color. *The Academy of Management Review, 22*(4), 946-973. https://doi.org/10.2307/259250

Gigone, D., & Hastie, R. (1997). Proper analysis of the accuracy of group judgments. *Psychological Bulletin, 121*(1), 149-167. https://doi.org/10.1037/0033-2909.121.1.149

Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review, 112*(2), 494-508. https://doi.org/10.1037/0033-295X.112.2.494494

Hauer, K. E., ten Cate, O., Boscardin, C. K., Iobst, W., Holmboe, E. S., Chesluk, B., . . . O'Sullivan, P. S. (2016). Ensuring resident competence: A narrative review of the literature on group decision making to inform the work of clinical competency committees. *Journal of Graduate Medical Education, 8*(2), 156-164. https://doi.org/10.4300/JGME-D-15-00144.1

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121*(1), 43-64. https://doi.org/10.1037/0033-2909.121.1.43

Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology, 50*(6), 1141-1151. https://doi.org/10.1037/0022-3514.50.6.1141

Jackson, S. E., May, K. E., & Whitney, K. (1995). Understanding the dynamics of diversity in decision-making teams. In R. A. Guzzo & E. Salas (Eds.), *Team*

effectiveness and decision making in organizations (pp. 204-261). San Franciso, CA: Jossey-Bass.

Janis, I. L. (1982). *Groupthink (2nd ed.)* Boston, MA: Houghton Mifflin.

Kameda, T., Tindale, R. S., & Davis, J. H. (Eds.). (2003). *Cognitions, preferences, and social sharedness: Past, present, and future directions in group decision making*. New York, NY: Cambridge University Press.

Kameda, T., Wisdom, T., Toyokawa, W., & Inukai, K. (2012). Is consensus-seeking unique to humans? A selective review of animal group decision-making and its implications for (human) social psychology. *Group Processes & Intergroup Relations, 15*(5), 673-689. https://doi.org/10.1177/1368430212451863

Kerr, N. L., Davis, J. H., Meek, D., & Rissman, A. K. (1975). Group position as a function of member attitudes: Choice shift effects from the perspective of social decision scheme theory. *Journal of Personality and Social Psychology, 31*(3), 574-593. https://doi.org/10.1037/h0076483

Kerr, N. L., MacCoun, R. J., & Kramer, G. P. (1996). Bias in judgment: Comparing individuals and groups. *Psychological Review, 103*(4), 687-719. https://doi.org/10.1037/0033-295X.103.4.687

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review of Psychology, 55*, 623-655. https://doi.org/10.1146/annurev.psych.55.090902.142009

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal*

of Applied Psychology, 85(2), 273-283. https://doi.org/10.1037/0021-9010.85.2.272

Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology, 94*(2), 535-546. https://doi.org/10.1037/a0013773

Miller, J. G. (1960). Information input overload and psychopathology. *The American Journal of Psychoanalysis, 116*(8), 695-704. https://doi.org/10.1176/ajp.116.8.695

Mohammed, S., & Dumville, B. C. (2001). Team mental models in a team knowledge framework: Expanding theory and measurement across disciplinary boundaries. *Journal of Organizational Behavior, 22*(2), 89-106. https://doi.org/10.1002/job.86

Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin, 83*(4), 602-627. https://doi.org/10.1037/0033-2909.83.4.602

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1), 1-13. https://doi.org/10.1177/1609406917733847

Odorizzi, S., Cheung, W. J., Sherbino, J., Lee, A. C., Thurgur, L., & Frank, J. R. (2020). A signal through the noise: Do professionalism concerns impact the decision-making of competence committees? *Academic Medicine, 95*(6), 896-901. https://doi.org/10.1097/ACM.0000000000003005

Pack, R., Lingard, L., Watling, C. J., Chahine, S., & Cristancho, S. M. (2019). Some assembly required: Tracing the interpretive work of clinical competency

committees. *Medical Education, 53*(7), 723-734. https://doi.org/10.1111/medu.13884

Pack, R., Lingard, L., Watling, C. J., & Cristancho, S. (2020). Beyond summative decision-making: Illuminating the broader roles of competence committees. *Medical Education, 54*(6), 517-527. https://doi.org/10.1111/MEDU.14072

Page, S. E. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and society*. Princeton, NJ: Princeton University Press.

Ramya, N., & Ali, S. A. M. (2016). Factors affecting consumer buying behavior. *International Journal of Applied Research 2*(10), 76-80.

Schumacher, D. J., Michelson, C., Poynter, S., Barnes, M. M., Li, S.-T. T., Burman, N., . . . Appd Learn CCC Study Group. (2018). Thresholds and interpretations: How clinical competency committees identify pediatric residents with performance concerns. *Medical Teacher, 40*(1), 70-79. https://doi.org/10.1080/0142159X.2017.1394576

Seeley, T. D. (2010). *Honeybee democracy*. Princeton, NJ: Princeton University Press.

Seeley, T. D., & Buhrman, S. C. (1999). Group decision makinig in swarms of honey bees. *Behavioral Ecology and Sociobiology, 45*(1), 19-31. https://doi.org/10.1007/s002650050536

Stasser, G. (1999). A primer of social decision scheme theory: Models of group influence, competitive model-testing, and prospective modeling. *Organizational Behavior and Human Decision Processes, 80*(1), 3-20. https://doi.org/10.1006/obhd.1999.2851

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48*(6), 1467-1478. https://doi.org/10.1037/0022-3514.48.6.1467

Tindale, R. S., & Kameda, T. (2000). 'Social sharedness' as a unifying theme for information processing in groups. *Group Processes & Intergroup Relations, 3*(2), 123-140. https://doi.org/10.1177/1368430200003002002

Tindale, R. S., & Sheffey, S. (2002). Shared information, cognitive load, and group memory. *Group Processes & Intergroup Relations, 5*(1), 5-18. https://doi.org/10.1177/1368430202005001535

Tindale, R. S., Smith, C. M., Dykema-Engblade, A., & Kluwe, K. (2012). Good and bad group performance: Same process—different outcomes. *Group Processes & Intergroup Relations, 15*(5), 603-618. https://doi.org/10.1177/1368430212454928

Von Cranach, M., Ochsenbein, G., & Valach, L. (1986). The group as a self-active system: Outline of a theory of group action. *European Journal of Social Psychology, 16*(3), 193-229. https://doi.org/10.1002/ejsp.2420160302

Zuber, J., Crott, H., & Wener, J. (1992). Choice shift and group polarization: An analysis of the status of arguments and social decision schemes. *Journal of Personality and Social Psychology, 62*(1), 50-61. https://doi.org/10.1037/0022-3514.62.1.50

Tables and Figures

Table 1. Novice raters: Comparison of decisions made by groups and decisions made by individuals in each condition using permutation tests

Condition	**Significance**
Control	$p = .08$
Strict assessors	$p < .01$*
Prior positive experience	$p < .01$*
Lenient assessors	$p < .01$*
Difficult personal circumstance	$p < .01$*
Unprofessional behaviour	$p = .02$*

* = Statistically significant at the $p < .05$ level

Table 2. Novice raters: Comparison of variance in decisions made by groups and decisions made by individuals in each condition using paired *t*-tests

Condition	**Test Statistic and Significance**
Control	$t(6) = -1.92$, $p = .10$
Strict assessors	$t(6) = -1.68$, $p = .14$
Prior positive experience	$t(6) = -4.72$, $p < .01$*
Lenient assessors	$t(6) = -4.15$, $p < .01$*
Difficult personal circumstance	$t(6) = -3.35$, $p = .02$*
Unprofessional behaviour	$t(6) = -1.36$, $p = .22$

* = Statistically significant at the $p < .05$ level

Table 3. Experienced raters: Comparison of decisions made by groups and decisions made by individuals in each condition using permutation tests

Condition	**Significance**
Control	$p < .01$*
Strict assessors	$p < .01$*
Prior positive experience	$p < .01$*
Lenient assessors	$p < .01$*
Difficult personal circumstance	$p < .01$*
Unprofessional behaviour	$p < .01$*

* = Statistically significant at the $p < .05$ level

Table 4. Experienced raters: Comparison of variance in decisions made by groups and decisions made by individuals in each condition using paired *t*-tests

Condition	**Test Statistic and Significance**
Control	$t(6) = 0.25$, $p = .81$
Strict assessors	$t(6) = 0.00$, $p = 1.00$
Prior positive experience	$t(6) = -1.97$, $p = .10$
Lenient assessors	$t(6) = -2.10$, $p = .08$
Difficult personal circumstance	$t(6) = -0.48$, $p = .65$
Unprofessional behaviour	$t(6) = -0.82$, $p = .44$

* = Statistically significant at the $p < .05$ level

Table 5. Examples of anecdotal evidence from analysis of CC meeting transcripts

Type of Anecdotal Evidence	Example
Strict assessor	*Chair:* "So, Dr. X gave two twos … uh, yeah … that's maybe a bit harsh … must be a hawk."
Prior positive experience	*Member:* "The first assessment was 'I needed to be there just in case' [four on the five-point EPA scale]." *Academic Advisor:* "It may have been easy though." *Program Director:* "I remember when [Resident] worked with Dr. X, [they] said that [Resident] is amazing. So maybe [Resident] caught onto everything." *Chair*: "That's good." *Program Director:* "[Dr. X] told me that [Resident] is amazing. … I didn't ask—[Dr. X] just said it." *Academic Advisor:* "It seems like [Resident] is competent then."
Lenient assessor	*Program Director:* "So, for all our PGY2s we don't have assessments from [rotation] … all's we know is that there's never been a PGY2 resident go through [rotation] where they aren't told that they're excellent and should consider a career in [subspecialty]. This resident received the same feedback."
Difficult personal circumstance	*Member 1:* "[Resident] has all of the right hands, right ideas, and so on. But [they're] a little bit lazy." *Member 2:* "Yeah, I noticed that when [Resident] came to [site]. [They] missed a lot of things." *Program Director:* "I have a bit of background on what is going on. Is it okay if I mention it? [Resident] is having a lot of problems in terms of [their] health…"
Unprofessional behaviour	*In one CC meeting, a lengthy discussion ensued in which numerous anecdotes were shared about a resident's unprofessional behaviour. These cannot be shared for reasons of confidentiality; however, the anecdotal nature of these comments is captured in the remarks below:* *Member:* "There's a documentation train right now that doesn't completely reflect the concerns about [Resident]. But there is strong anecdotal evidence and strong additional information has come to light that makes it quite clear that [they're] not ready to be promoted."

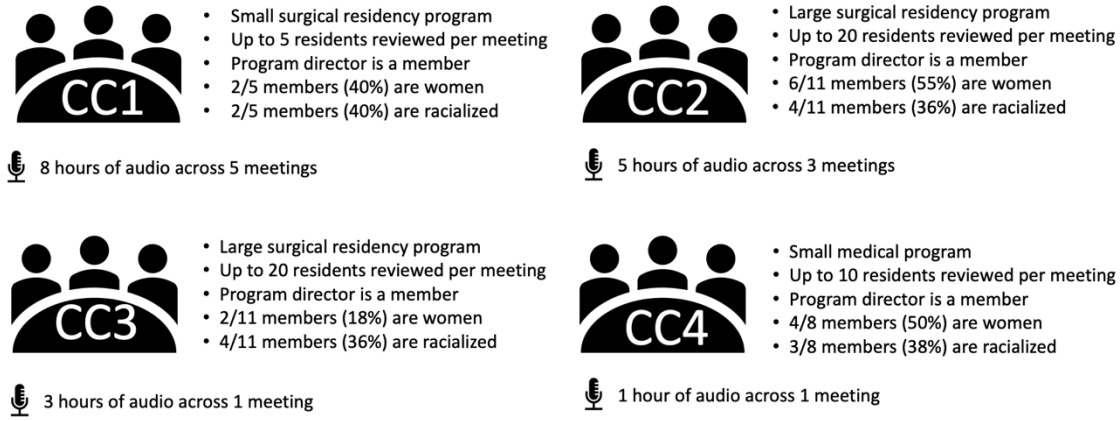

Figure 1. Characteristics of the four CCs from which meeting transcripts were obtained, along with the amount of data available for each

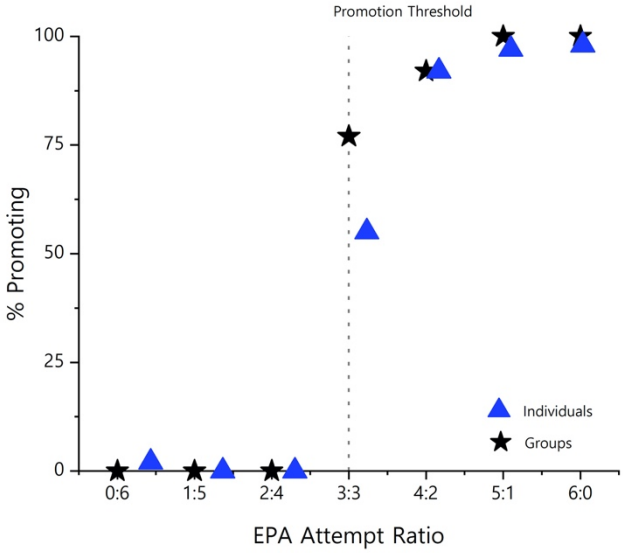

Figure 2a. Novice raters: Comparison of individual (blue) and group (black) promotion decision curves for the control condition

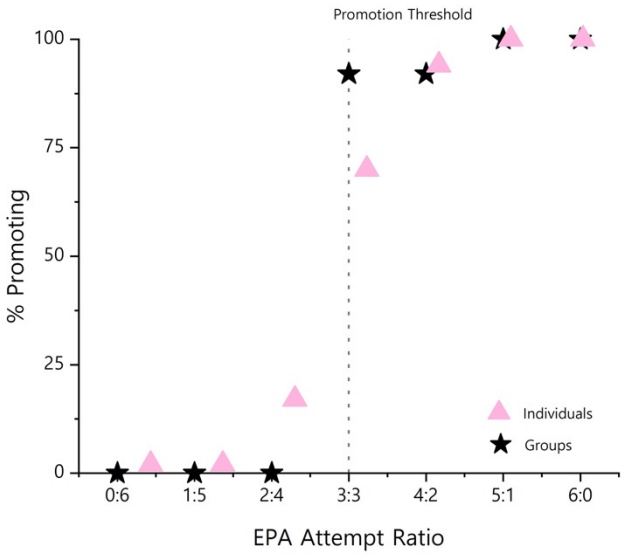

Figure 2b. Novice raters: Comparison of individual (pink) and group (black) promotion decision curves for the strict assessors condition

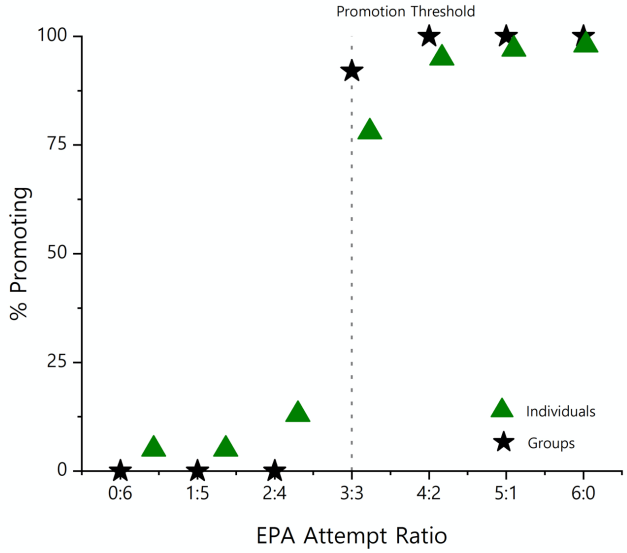

Figure 2c. Novice raters: Comparison of individual (green) and group (black) promotion decision curves for the prior positive experience condition

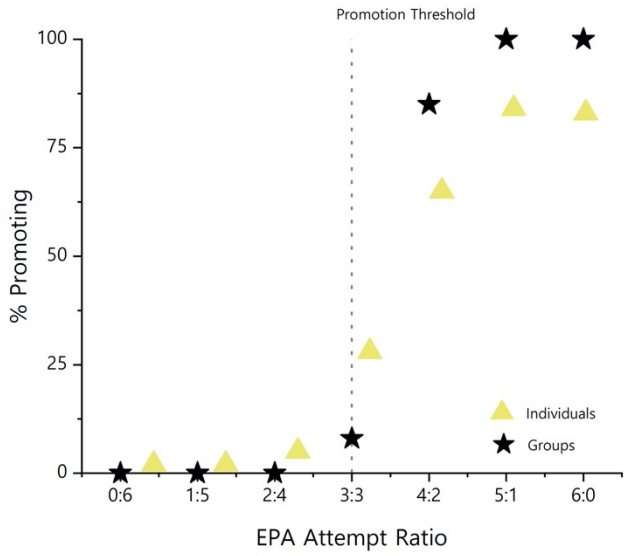

Figure 2d. Novice raters: Comparison of individual (yellow) and group (black) promotion decision curves for the lenient assessors condition

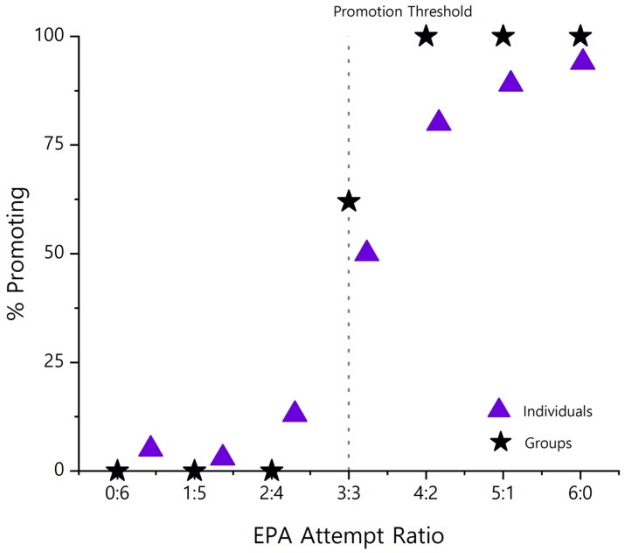

Figure 2e. Novice raters: Comparison of individual (purple) and group (black) promotion decision curves for the difficult personal circumstance condition

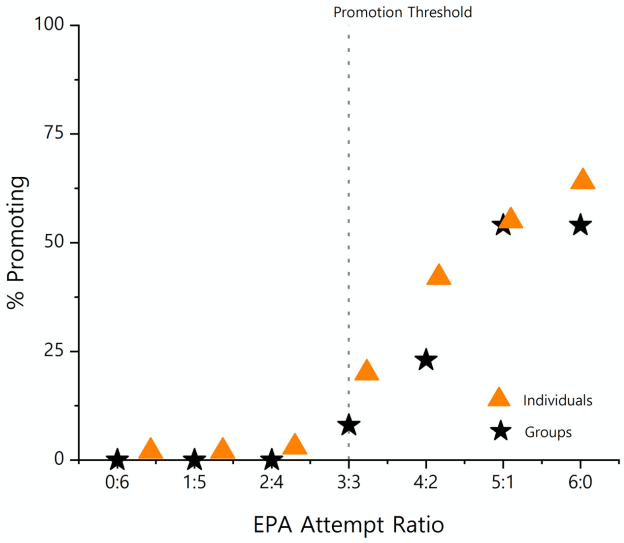

Figure 2f. Novice raters: Comparison of individual (orange) and group (black) promotion decision curves for the unprofessional behaviour condition

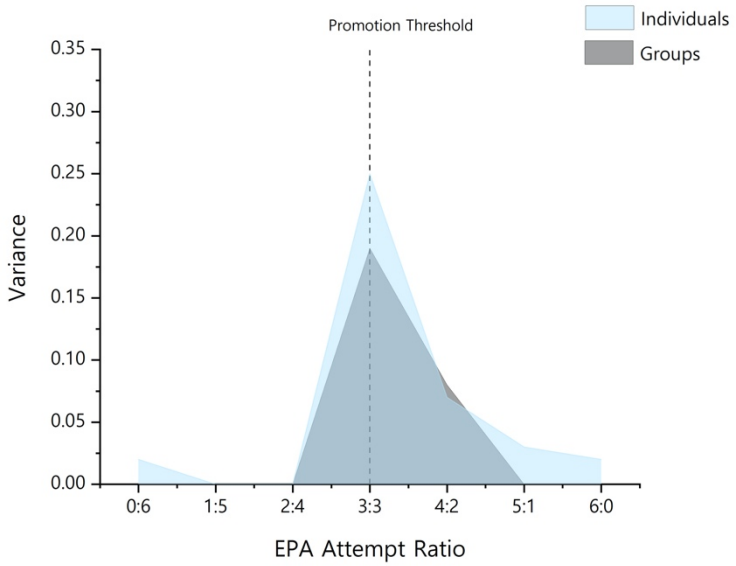

Figure 3a. Novice raters: Variance in promotion decisions for decisions made individually (blue) and in groups (grey) in the control condition

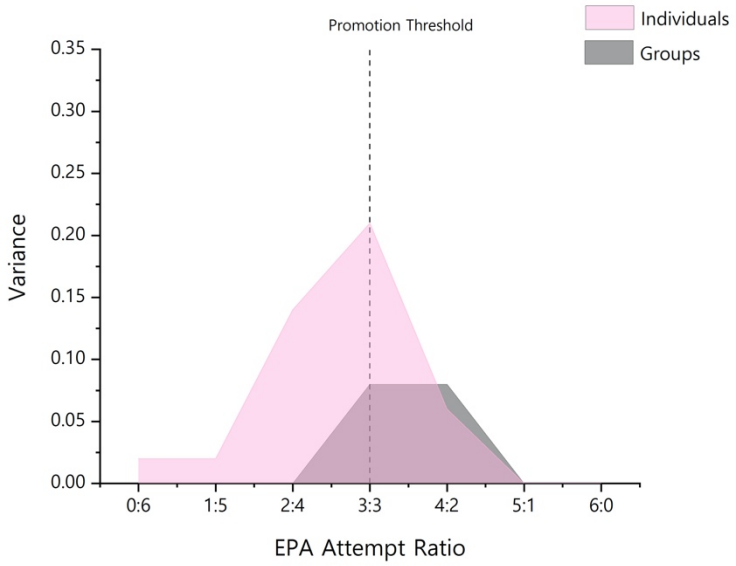

Figure 3b. Novice raters: Variance in promotion decisions for decisions made individually (pink) and in groups (grey) in the strict assessors condition

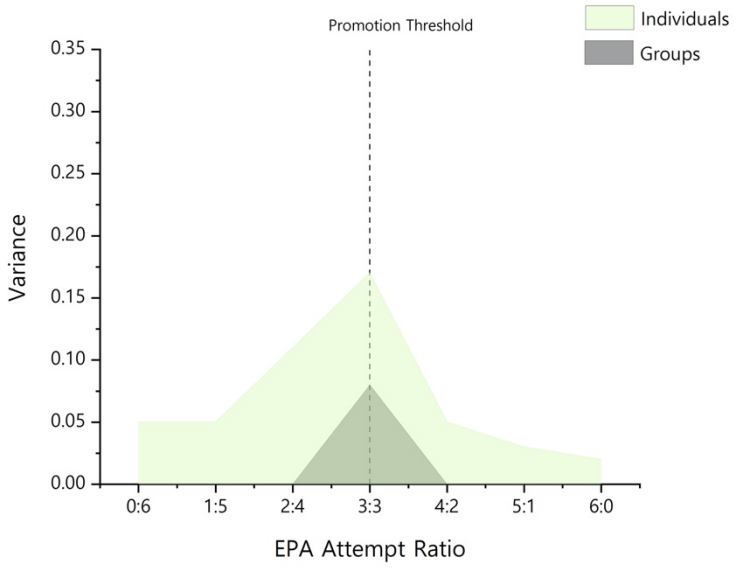

Figure 3c. Novice raters: Variance in promotion decisions for decisions made individually (green) and in groups (grey) in the prior positive experience condition

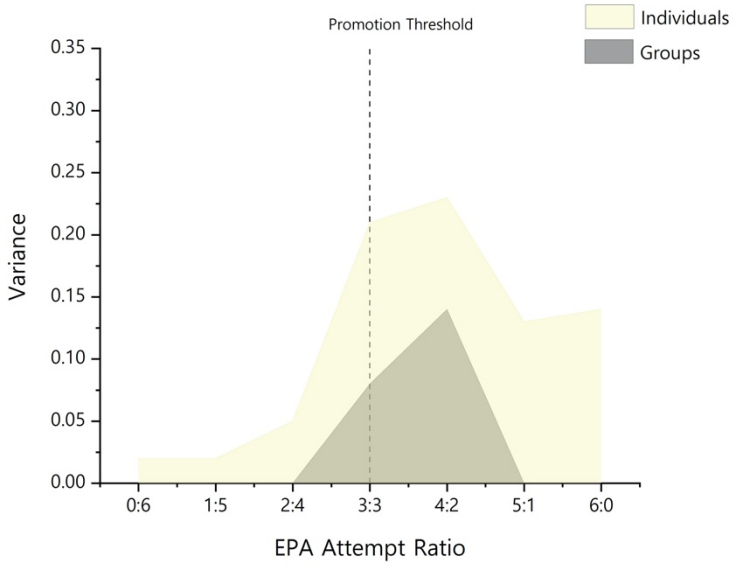

Figure 3d. Novice raters: Variance in promotion decisions for decisions made individually (yellow) and in groups (grey) in the lenient assessors condition

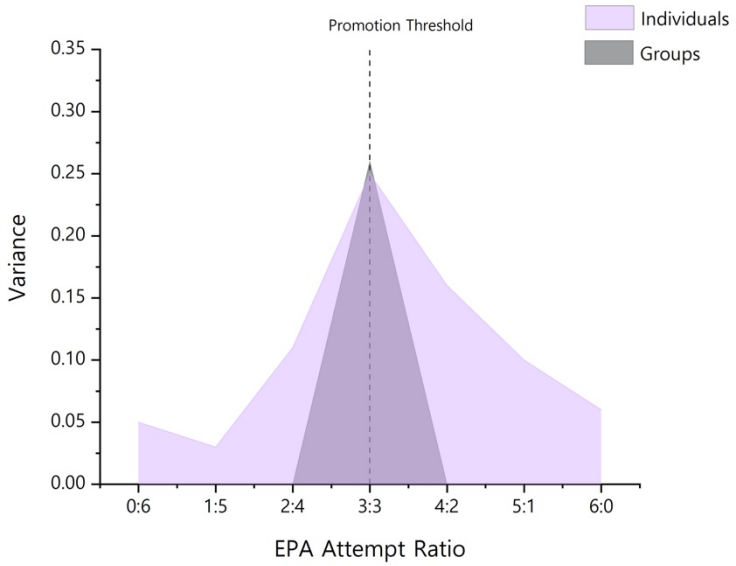

Figure 3e. Novice raters: Variance in promotion decisions for decisions made individually (purple) and in groups (grey) in the difficult personal circumstance condition

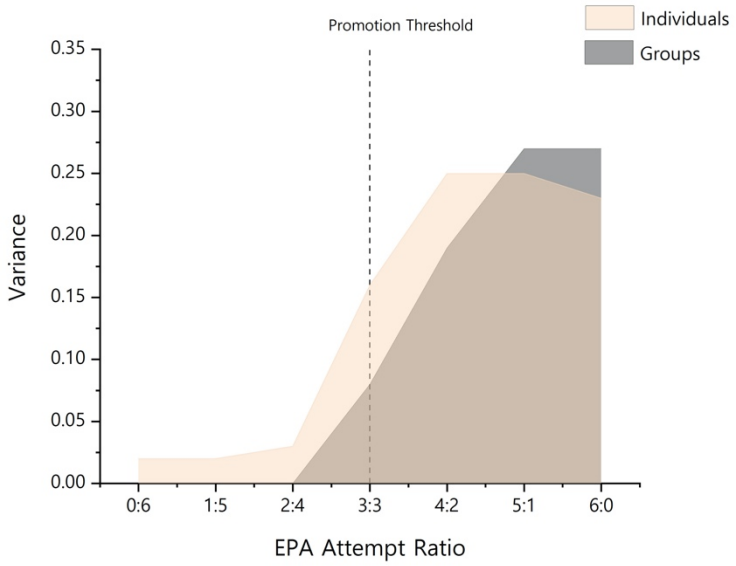

Figure 3f. Novice raters: Variance in promotion decisions for decisions made individually (orange) and in groups (grey) in the unprofessional behaviour condition

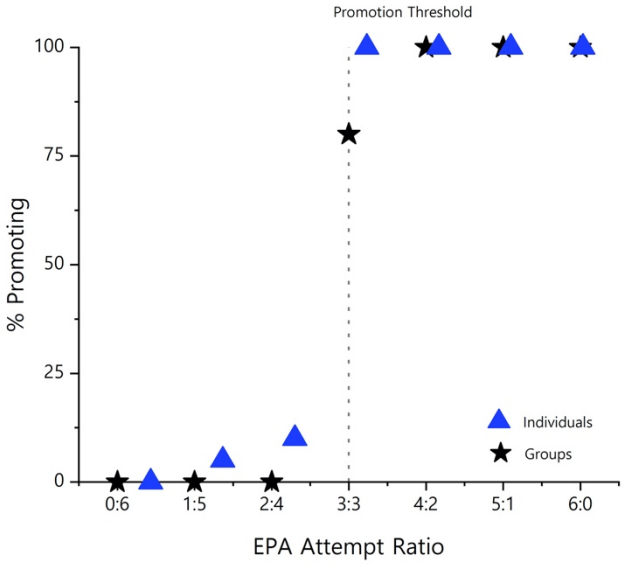

Figure 4a. Experienced raters: Comparison of individual (blue) and group (black) promotion decision curves for the control condition

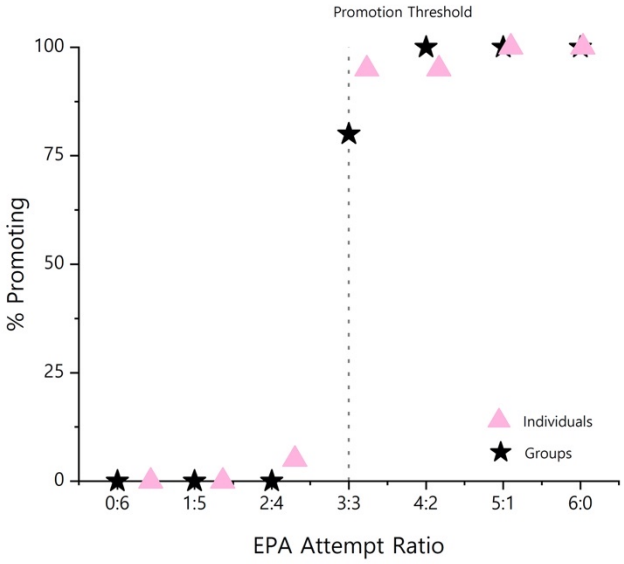

Figure 4b. Experienced raters: Comparison of individual (pink) and group (black) promotion decision curves for the strict assessors condition

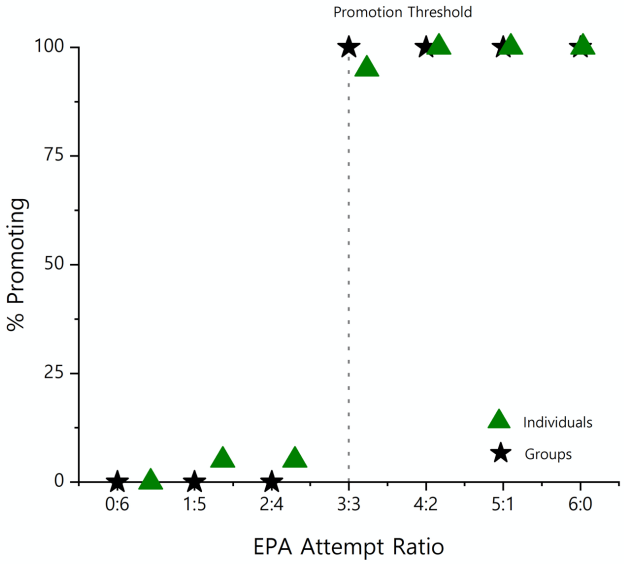

Figure 4c. Novice raters: Comparison of individual (green) and group (black) promotion decision curves for the prior positive experience condition

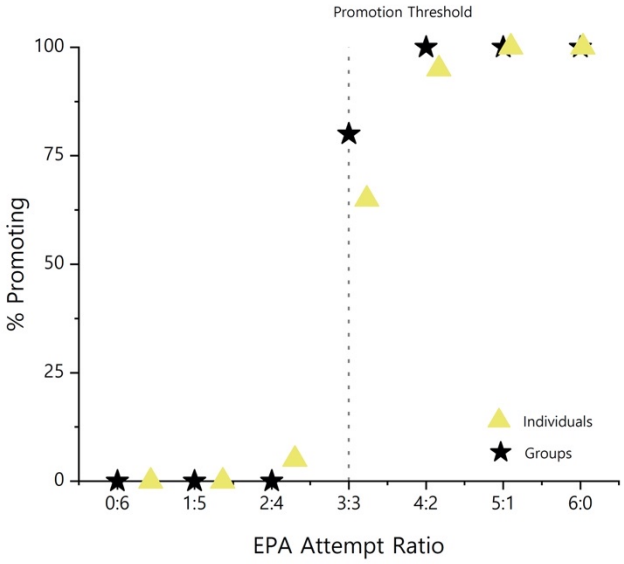

Figure 4d. Experienced raters: Comparison of individual (yellow) and group (black) promotion decision curves for the lenient assessors condition

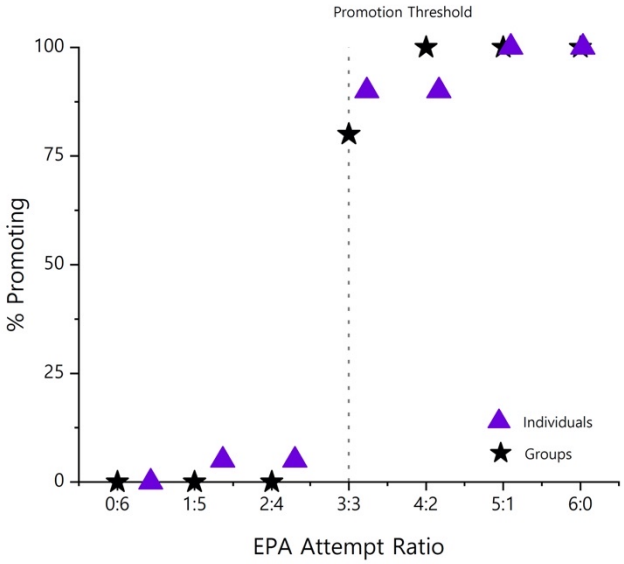

Figure 4e. Experienced raters: Comparison of individual (purple) and group (black) promotion decision curves for the difficult personal circumstance condition

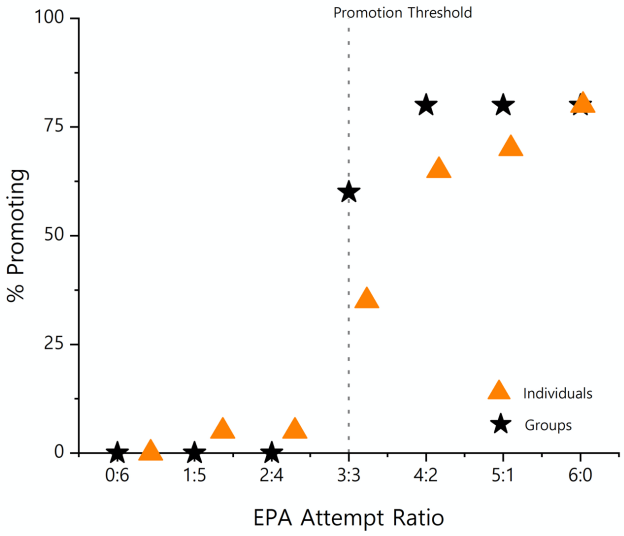

Figure 4f. Experienced raters: Comparison of individual (orange) and group (black) promotion decision curves for the unprofessional behaviour condition

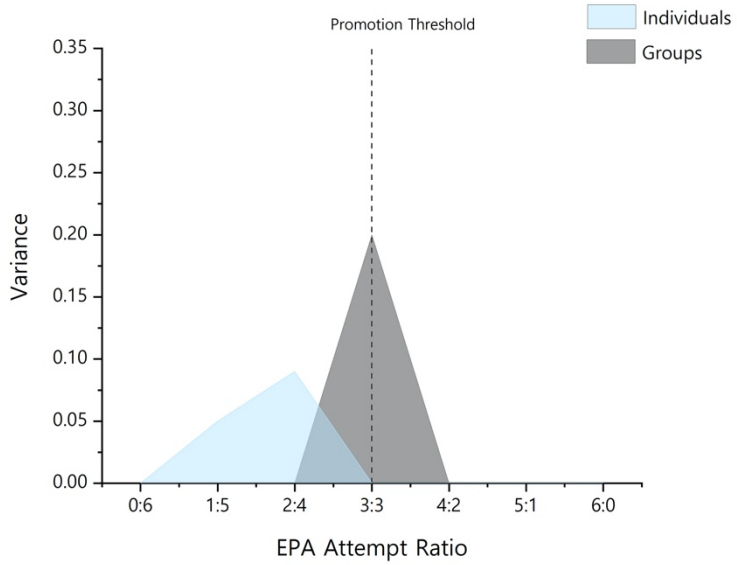

Figure 5a. Experienced raters: Variance in promotion decisions for decisions made individually (blue) and in groups (grey) in the control condition

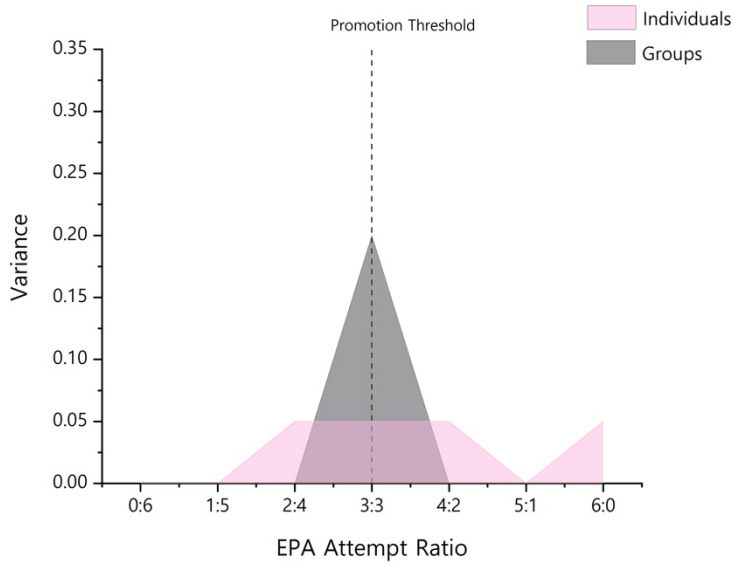

Figure 5b. Experienced raters: Variance in promotion decisions for decisions made individually (pink) and in groups (grey) in the strict assessors condition

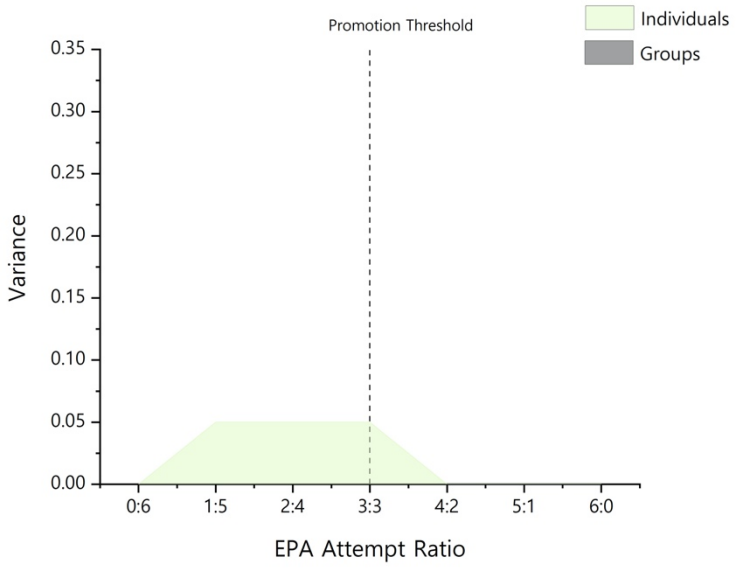

Figure 5c. Experienced raters: Variance in promotion decisions for decisions made individually (green) and in groups (grey) in the prior positive experience condition

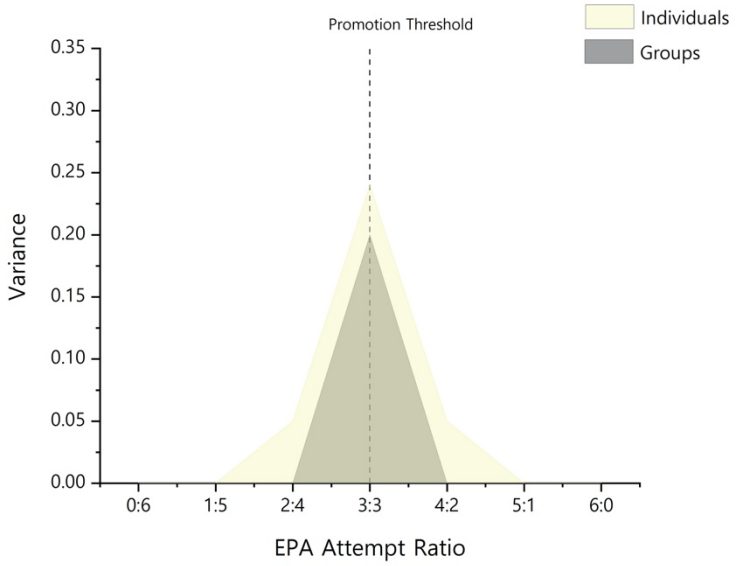

Figure 5d. Experienced raters: Variance in promotion decisions for decisions made individually (yellow) and in groups (grey) in the lenient assessors condition

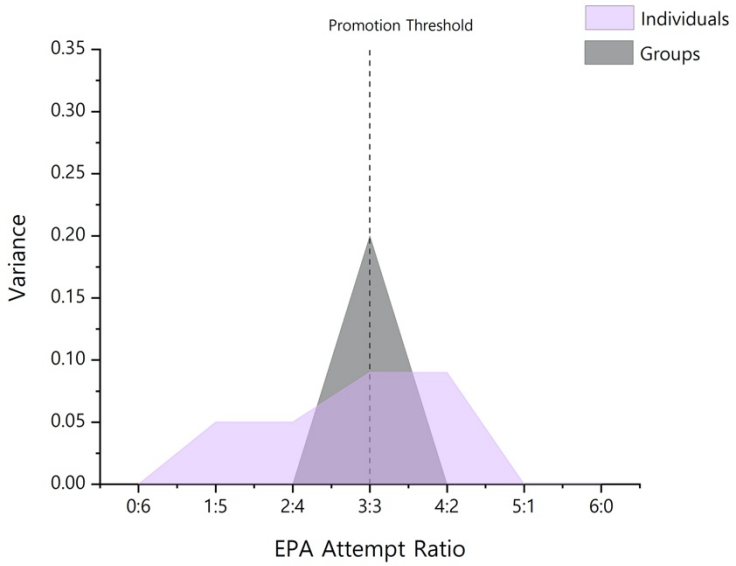

Figure 5e. Experienced raters: Variance in promotion decisions for decisions made individually (purple) and in groups (grey) in the difficult personal circumstance condition

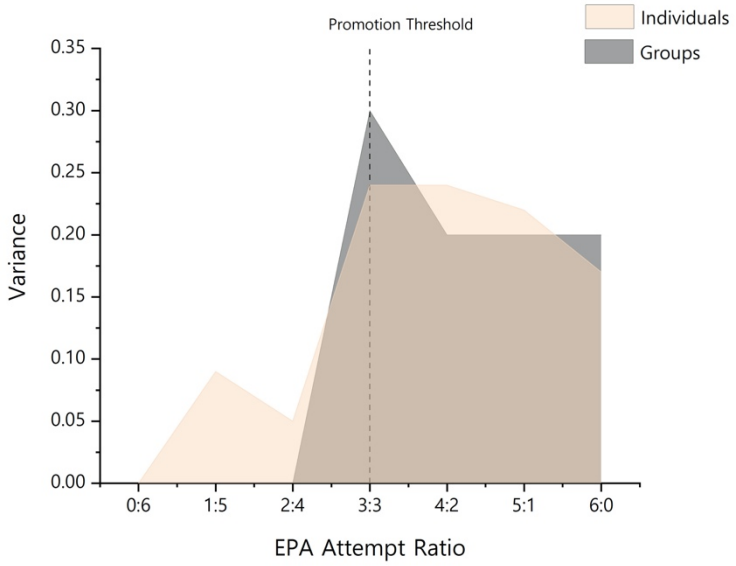

Figure 5f. Experienced raters: Variance in promotion decisions for decisions made individually (purple) and in groups (grey) in the difficult personal circumstance condition

Chapter 5

The Role of Social Hierarchies and Power in Competence Committee Decision Making

5.1 Introduction

Information processing in groups depends considerably on how power is distributed among group members (Hinsz, Tindale, & Vollrath, 1997). Certain social hierarchies may facilitate open discussion while others may inhibit it. For example, groups in which power is held by one or two dominant members may have poorer information-sharing practices than groups in which power is more equally distributed among members (Hinsz et al., 1997). This can manifest in different ways: members may share fewer ideas or none at all, or they may conform to more dominant group members' opinions during the group decision-making process (Elsass & Graves, 1997; Hinsz et al., 1997). Thus, in order to fully understand how CCs make decisions, researchers must understand how their members experience, and as a result, contribute to the group decision-making process.

5.1.1 Understanding Members' Experiences and Participation in Groups

Elsass and Graves (1997) proposed a theoretical framework for understanding how various cognitive and behavioural factors may influence members' experiences and participation in groups. It is noteworthy that these factors can, and frequently do, evolve over time through feedback processes. The first phase of the model depicts a categorization process through which members engage in social comparison to determine role expectations. These categorization processes are automatic and based on members'

social identities, which consist of their social category (e.g., gender, race/ethnicity, profession) and personal identity (e.g., personality or attributes; Elsass & Graves, 1997).

Evidence for automatic categorization processes among group members comes from both the cognitive and neuroscientific literature. For example, both adults and children spend more time looking at and attending to high-status individuals than low-status individuals (Abramovitch & Grusec, 1978; Foulsham, Cheng, Tracy, Henrich, & Kingstone, 2010). This attentional bias also has an effect on memory, such that people tend to be quicker and more accurate at recognizing high-status faces than low-status faces (Ratcliff, Hugenberg, Shriver, & Bernstein, 2011). Moreover, parts of the brain's emotional response and reward systems are activated when individuals pay attention to high-status individuals, suggesting that doing so may be adaptive and confer a benefit to survival (e.g., by ensuring access to resources; Delgado, 2007).

The way in which individuals are categorized by themselves and others leads to different role expectations when working in groups (Elsass & Graves, 1997). While high-status individuals may develop role expectations that are more focused on leadership and power, low-status individuals' role expectations can be negatively influenced by stereotypes, leading them to take on a diminished role within the group. The second phase of Elsass and Graves' model describes the behavioural interactions between group members on the basis of their varied statuses within the group. If role expectations for a particular group member are high, then they are more likely to take on a more dominant role within the group. Their actions—even if they break the rules or behave unethically—are generally perceived as more acceptable than those of low-status individuals (Piff,

Stancato, Côté, Mendoza-Denton, & Keltner, 2012). Conversely, if expectations of a particular group member are low, then that member is likely to be treated as being of lower status within the group, such as by being assigned tasks of lower importance or having their contributions devalued (Elsass & Graves, 1997). Importantly, members' behaviours and group members' reactions can perpetuate low role expectations for lower status members by reinforcing preliminary negative expectations. This may explain, for example, why women and people of colour tend to speak less in group exchanges than men and white-status individuals (Elsass & Graves, 1997).

The final phase of the model elucidates how social and cognitive processes interact to produce different outcomes and patterns of interaction within groups. Specific patterns include engagement, contributing, complementing, and exclusion (Elsass & Graves, 1997). Engagement stems from high quality instrumental and social exchanges and indicates a high-functioning group in which low role expectations for the outgroup member have been avoided, and they are an active contributor to the group. Contributing refers to high instrumental but low social exchange, leading a group member to contribute to the group's task but feel socially isolated while doing so. Complementing is the opposite pattern—that is, low instrumental but high social exchange. This pattern is characterized by an outgroup member who engages in social interactions with group members but does not meaningfully contribute to the group's task. Finally, exclusion refers to low instrumental and social exchange, wherein the member is fully excluded from both social and task-related processes within the group (Elsass & Graves, 1997).

5.1.2 Bases of Power

Another theoretical framework for understanding social influences in groups is French and Raven's (1959) bases of power. According to the authors, social power is the ability of one person's actions to alter another person's beliefs, attitudes, or behaviours. This can occur in different ways, as per the five 'bases,' or forms, of power: coercive, reward, legitimate, referent, and expert. A sixth base, informational, was added some years after the development of the original framework (Raven, 1965).

Coercive power occurs when the influencing agent uses force or threat of force to persuade the other party (French & Raven, 1959). Force can constitute different forms, including physical, social, emotional, political or economic, and does not have to be obvious to the target. Reward power occurs when the influencing agent offers a reward to incentivize the target to adopt their perspective (French & Raven, 1959). Rewards can be impersonal (e.g., money) or personal (e.g., approval; Raven, 1992). Furthermore, since social belonging is linked to social sharedness, people may feel compelled to behave in ways that align with the values held by the groups and organizations to which they belong—this is known as referent power (French & Raven, 1959).

Legitimate power indicates that the influencing agent holds power by virtue of being in a position of authority (e.g., due to being in an elected or appointed role; French & Raven, 1959). Social norms generally dictate that people in positions of authority be respected, thereby making them more influential than other members when working in a group setting. A related idea is expert power, which is influence through one's knowledge, experience, skills, or talents (French & Raven, 1959). Like referent power,

both legitimate and expert power can result in both positive and negative outcomes. Finally, informational power is linked to influence through the information that one possesses and is able to share with others (Raven, 1965).

5.1.3 Social Hierarchies and Power in Competence Committee Decision Making

As previously mentioned, the various social categories to which an individual belongs can influence their experiences in group settings. Some prominent examples of social categories include, but are not limited to, gender, race/ethnicity, position, and socioeconomic status. Importantly, the various dimensions of a person's social identity can interact with each other to shape one's experiences in social settings, including their power and privilege (Crenshaw, 1989). This idea, known as 'intersectionality,' originated from a critical analysis of Black women's experiences in the legal system but has since been expanded to include many other dimensions of identity, including ability, age, religion, sexual orientation, and more (Crenshaw, 1989; Mays & Ghavami, 2018).

Certain social factors may be more relevant to CCs than others. Since medicine is a hierarchical profession (Farnan, Johnson, Meltzer, Humphrey, & Arora, 2008; Vanstone & Grierson, 2019), it is possible that certain members of a CC, such as senior faculty, play a more dominant role on the committee than junior members. This could occur through legitimate, expert, and/or informational power (French & Raven, 1959). An impact of referent power is also possible if some members feel compelled to act in a way that aligns with the committee's norms. This would not only depend on the extent to which the committee has established norms, but also how strongly members identify with their role on the committee (Brewer, 1991; Hinsz et al., 1997).

Gender may impact CC decision-making processes in that the gender distribution of faculty in certain programs may be heavily skewed towards one gender over the other (e.g., surgical fields often have more men than women whereas certain medical fields have a more equal distribution of women and men; Abelson, Chartrand, Moo, Moore, & Yeo, 2016), and any systematic differences would likely be reflected in the committee's membership also. A common finding in the social psychological literature is that men tend to speak more, and more often, than women in group settings (Elsass & Graves, 1997; Kelly, Wildman, & Urey, 1982; Kirchmeyer, 1996). Thus, it is possible that similar differences exist in the context of CCs, as well, such that women may speak less than men and, as a result, hold less decision-making power. This effect may be exacerbated in committees whose membership is imbalanced and skewed towards men. On the other hand, women that persist in surgical careers may be less likely to conform to conventional stereotypes (Moulton, Seemann, & Webster, 2013; Webster et al., 2016); thus, it is also plausible that fewer gender differences exist within CCs than elsewhere. These different possibilities suggest a need to further investigate this topic.

Race/ethnicity may impact CC decision-making processes as racialized members tend to speak less often than white members in group settings (Elsass & Graves, 1997; Li, Karakowsky, & Siegel, 1999; Sommers, 2006). This is particularly true if they constitute a minority of the group's membership or have other characteristics that may intersect with their race/ethnicity to result in even greater marginalization—for example, being a woman or being junior (Crenshaw, 1989; Elsass & Graves, 1997). On the other hand, as for women, racialized members who have 'made it' in the medical profession may be less

likely to conform to traditional stereotypes than those in other contexts, once again underscoring the need to investigate this topic in further detail.

The relative socioeconomic status of members is also a consideration when exploring the role of social processes in groups (Sheehy-Skeffington, 2020). However, this factor is likely less relevant in the context of CCs because the majority of members would be of a similar (i.e., high) socioeconomic status by virtue of being in the medical profession. Thus, this factor will not be considered further for the purposes of this thesis.

5.1.4 Study Purpose and Overview

This chapter will examine the role of social hierarchies and power in CC decision making, and how factors such as members' position, gender, and race/ethnicity affect their interactions within the committee. Although these factors do not, in any way, encompass all of the social categories to which CC members may belong, they are the focus of this chapter because of their visibility and salience in social contexts (Stewart & Valian, 2018). First, the number of speaking turns and words spoken by each committee member will be analyzed as a function of position, gender, and race/ethnicity across four CCs. Next, the personal characteristics of individuals contributing anecdotal evidence during CC meetings will be studied, along with their influence on the decisions made by the committee. Finally, qualitative interviews will be used to study members' lived experiences of social hierarchies and power in CCs.

5.2 Methods

5.2.1 Number of Speaking Turns and Words Spoken

Transcripts from the four CCs that gave permission to have their meetings audio recorded in Chapter 4 were used for analysis in this study. Their characteristics and the amount of data available for each are presented in Figure 1.

In social psychological studies, the number of speaking turns and words spoken are often used as proxies for social influence (Brooke & Ng, 1986; Ng & Bradac, 1993). It is assumed that group members who speak more are more likely to be able to influence the committee's decisions than individuals who speak less or not at all (Brooke & Ng, 1986; Ng & Bradac, 1993). In keeping with this idea, each transcript included in the present study was reviewed and the number of speaking turns and words spoken were extracted for each member. Values for both variables were averaged across meetings of the same committee and stratified by position, gender, and race/ethnicity.

5.2.2 Promotion Decisions

Although the number of speaking turns and words spoken per member provide some insight into which CC members are most influential, this method is not foolproof as it does not account for the fact that some members may speak infrequently but still make impactful contributions (i.e., quality versus quantity). Thus, another way of studying the role of social hierarchies and power in CCs is to examine the extent to which information shared by a particular member impacts the committee's decision making. Since the primary focus of this thesis was on anecdotal evidence, analyses were limited to when this type of information was shared by a committee member.

Transcripts from the four CCs included in the study were analyzed to identify instances in which the committee had made a promotion decision. For each decision, the following information was recorded in a table: the number of promotion decisions made, the number involving anecdotal evidence, and the number in which the outcome appeared to change as a result of the anecdotal evidence provided. In cases where anecdotal evidence appeared to alter the committee's decision making, the positions of all members providing anecdotal evidence were recorded to determine which members were the most influential (i.e., chair, program director, other senior member, regular member, junior member, or academic advisor). Junior members were residents or faculty with five or fewer years of experience in their role, regular members were faculty with six to nine years of experience in their role, and other senior members were faculty with ten or more years of experience in their role who were not the chair or program director.

In conducting these analyses, the same definition was used for anecdotal evidence as in earlier chapters of this thesis: undocumented information shared by a CC member about a resident or their assessors. Examples included members' personal knowledge about an assessor's standards or a resident's behaviours, skills, or personal circumstances. Whether or not an anecdote appeared to alter the committee's decision making was a subjective determination made by the primary investigator; however, the deciding factor was whether the committee appeared to be leaning in a different direction before learning the information. A common example was when a committee would have held a resident back on the basis of missing assessment data were it not for anecdotes shared by one or more members that demonstrated the resident's competence on the skill(s) in question.

5.2.3 Competence Committee Member Perceptions and Experiences

The first two analyses examined social hierarchies and power within CCs indirectly, through meeting transcripts. To more directly examine CC members' perceptions and experiences, a series of questions were included alongside the interview prompts used in Chapter 4. These questions probed whether members felt as if they could speak up when they had something to say, whether they felt that their contributions were heard, and whether their contributions were influenced by other committee members such as the chair, program director, or other senior committee members. The interviews also explored CC members' perceptions of each of these roles and what, if any, value they brought to the committee. Interview transcripts were analyzed using the procedures previously described in Chapter 4. Once again, guidelines from Nowell, Norris, White, and Moules (2017) were used to ensure the trustworthiness of the qualitative data.

5.3 Results

5.3.1 Number of Speaking Turns and Words Spoken

Figures 2 and 3 display the average number of speaking turns and words spoken per meeting, respectively, by various committee members. Across all committees, the program director and the chair spoke more than all other committee members, both in terms of the number of speaking turns and the number of words spoken. Program directors spoke the most, with an average of 171 speaking turns (*SD* = 94) and 5,292 words (*SD* = 4,072) per meeting. Chairs spoke the second most, with an average of 128 speaking turns (*SD* = 62) and 2,320 words (*SD* = 630) per meeting. The average for all other members was 36 speaking turns (*SD* = 16) and 777 words spoken (*SD* = 314) per

meeting. It is noteworthy that in two of the four CCs (CC1 and CC3), the program administrator had as many speaking turns as the average member.

Figures 4 and 5 show the number of speaking turns and words spoken per meeting, respectively, stratified by gender. Statistical tests were not conducted due to the small number of members per committee; thus, the following results indicate data trends with no implication of statistical significance. Visual inspection of the data, including error bars representing standard errors of the mean, suggested differences in two of the four committees. In CC2, women had a similar number of speaking turns as men but spoke fewer words, while in CC3, they had fewer speaking turns and spoke fewer words than men. Across all four committees, the average number of speaking turns per meeting was 65 (*SD* = 77) for women and 54 (*SD* = 36) for men, while the average number of words spoken per meeting was 1,292 (*SD* = 918) for women and 1,624 (*SD* = 554) for men.

A secondary analysis was conducted in which the chair and program director were removed from the analyses due to their influence on the committee described earlier. A senior member of CC2 was also removed as their contributions were more than two standard deviations above committee members not in leadership roles. When the chair and program director were removed, findings indicated that women spoke as much (CC2 and CC3) or even more than (CC1 and CC4) men (Figures 4 and 5). Across all four committees, the average number of speaking turns per meeting was 40 (*SD* = 28) for women and 19 (*SD* = 7) for men, while the average number of words spoken per meeting was 736 (*SD* = 205) for women and 541 (*SD* = 339) for men. The relatively large

standard deviations, particularly for women, indicate that women's contributions to CCs were not necessarily consistent; some contributed more than others.

Figures 6 and 7 show the number of speaking turns and words spoken per meeting, respectively, stratified by race/ethnicity. Visual inspection of the data, including error bars representing standard errors of the mean, suggested differences in only CC1, where racialized members appeared to have fewer speaking turns and speak fewer words than white members. Across all four committees, the average number of speaking turns per meeting was 54 (*SD* = 27) for racialized members and 74 (*SD* = 67) for white members, while the average number of words spoken per meeting was 1,197 (*SD* = 619) for racialized members and 1,755 (*SD* = 831) for white members. When the program director and chair were removed from the analyses due to their disproportionate influence, racialized members spoke as much as (CC1, CC2, and CC4) or even more than (CC3) white members (Figures 6 and 7). Across all four committees, the average number of speaking turns per meeting was 37 (*SD* = 19) for racialized members and 23 (*SD* = 8) for white members, while the average number of words spoken per meeting was 858 (*SD* = 495) for racialized members and 595 (*SD* = 214) for white members.

5.3.2 Promotion Decisions

Table 1 shows the number of promotion decisions made by each CC, the number involving anecdotal evidence, and the number in which the outcome appeared to change as a result of the anecdotal evidence provided. A total of 93 promotion decisions were made across the four CCs included in the analysis. The percentage of promotion decisions involving anecdotal evidence ranged from 15% to 86% across individual committees, for

an average of 31% (n = 29) across all committees. The number of times anecdotal evidence appeared to alter the decision outcome also varied by committee, ranging from 9% to 75%, for an average of 24% (n = 7) across all CCs. In all but one case, anecdotal evidence caused the committee to become stricter by changing their decision to promote or conditionally promote a resident to not promoting them.

Table 2 shows which committee members provided influential anecdotal evidence, or in other words, evidence that appeared to alter the committee's decision making. Of the 27 anecdotes captured, the majority were provided by program directors (33%; n = 9), followed by regular members (26%; n = 7) and chairs (22%; n = 6). It is noteworthy that a senior member of CC2 was a program director in another, related specialty. In conducting this analysis, this member was classified as a program director rather than a senior member because of their in-depth knowledge of the residents being reviewed. Thus, it appears that program directors may be influential not only on their home specialty's CC, but also committees on which they are serving as external members.

5.3.3 Competence Committee Member Perceptions and Experiences

Twelve CC members (three women and nine men) were interviewed. The sample represented CCs across seven specialties and a range of membership types (clinical faculty: n = 7; non-clinical members: n = 1; and residents: n = 4). Participants ranged in experience level from residents and junior faculty with less than five years of experience (n = 6) to senior faculty (n = 6), including program directors and CC chairs.

Overall, participants had very positive perceptions of the CC decision-making environment. Not only did CC members feel as if their opinions were being heard, but

they also felt that their contributions mattered: “I really appreciated … feeling like my voice was not only being heard but was valued and welcome” (Resident). Residents and junior faculty did not hold back when sharing their opinions with the committee: “I feel very comfortable speaking my mind to the faculty on this committee and they’re receptive to it. I never really feel like I need to hold back anything” (Resident). This was the case even when their opinions differed from that of other, more senior members: “Even if I feel like the staff are saying something that’s incorrect … or if they’re being unfair and I know something different … I’ll still speak up” (Resident).

In large part, these positive perceptions had to do with the fact that CC members were carefully selected to ensure a collegial environment: “The people I’m working with are very nice and open-minded and everything” (Junior Faculty). This was easier to do in smaller programs, in which faculty and residents were more likely to have established relationships with one another before becoming CC members: “It makes it a lot easier when you’re such a small program, to have that respect and that love for each other” (Resident). There were also elements of the CC decision-making process that members felt were conducive to encouraging input from all members. One example was assigning specific members to review each resident file before presenting a recommendation to the committee: “People are allotted time to speak as they’re reviewing residents, so that ability to discuss the residents they have looked at more closely … allows them to ask pertinent questions of the other reviewers” (CC Chair). Another was the data-driven nature of the CC review process: “Because it’s data-driven, I think people are a lot more happy to look at the data and then give their opinion” (CC Chair).

The interviews also probed members' perceptions of various roles on the committee, such as that of the chair and the program director. While members considered these roles important from an operational standpoint, most felt that their influence was equal to that of other members: "They have a lot of influence in how the committee is run, and some of the operations of it. … There's an equal say in determining whether or not somebody has met the criteria and should be moved to the next level" (Senior Faculty). However, there were a couple of exceptions to this viewpoint. One member suggested that certain individuals, such as the program director, had more influence on the CC because of the information they possessed: "I think people with defined leadership positions have an impact on how the discussion unfolds because the experience they bring to the discussion changes the input" (Program Director). Another member stated that these individuals could enact their influence in a subtle manner that was more difficult to recognize, such as by taking action outside of CC meetings that other members were not aware of: "I've seen things be tabled for next meeting, but then next meeting, things have happened behind the scenes that I wasn't necessarily aware of" (Resident).

While the chair was not perceived to be as influential as the program director in terms of the information they shared, participants recognized their importance in ensuring that committee meetings stayed on track and that all members had the opportunity to participate: "They gave me the floor … and gave me the opportunity to speak and feel safe" (Resident). In addition to encouraging all members to share their opinions, the chair also needed to be aware of when certain members were speaking too much: "The more difficult part is to try to not curtail but try to be aware of one or two people that tend to

dominate the discussion" (CC Chair). For this reason, members felt that the chair should be an experienced member with leadership training: "If you've got a senior leader there, then there's a power dynamic that the chair might have to learn how to manage and … if we really want them to do a good job, then they need some training" (Program Director).

5.4 Discussion

This chapter explored the role of social hierarchies and power in CCs. An analysis of how the number of speaking turns and words spoken varies by members' social status revealed that while gender and race/ethnicity did not appear to have a large influence on members' contributions, their position on the committee did. Program directors and chairs spoke the most out of any committee member, both in terms of the number of speaking turns and the number of words spoken. These results were corroborated by the findings of a second analysis, which showed that anecdotal evidence provided by program directors had the most influence on committees' decisions. Interviews with CC members revealed very positive perceptions about the extent to which members could speak up and have their voices heard. This was facilitated by features such as the careful selection of members, a prior history of working together, effective leadership, a shared purpose, and specific decision-making procedures such as review assignments and data-driven discussion. Interestingly, most members did not believe that senior members such as the program director or chair had undue influence on the committee's decisions.

5.4.1 Position on the Committee

Certain committee members, such as the program director and the chair, were highly influential. Not only did program directors speak more than any other committee

member, but they were also the most likely to share influential anecdotal evidence. This suggests that these individuals may be an important source of contextual information for CCs, which is consistent with their in-depth knowledge of residents in the program. This may especially be the case when a CC is still new, as it may not yet have established data sharing mechanisms that enable other committee members to access all components of a resident's file. Until such mechanisms are established, the program director may be the only member with access to certain information that is critical to the committee's decision making. A related implication is that since a substantial amount of program directors' time appeared to be spent sharing previously documented information with committee members, establishing mechanisms that allow information to be shared with members beforehand may lead to greater efficiency during meetings.

While chairs spoke the second most often out of any committee member, they were less likely to share influential anecdotal evidence with the committee than program directors. This suggests that their mechanism of influence may be different. Specifically, the chair may serve as a modulator of information flow through the CC by ensuring that the meeting stays on track and that each member is given an opportunity to speak. This is consistent with observational data collected throughout this thesis, which suggested that chairs rarely held strong personal convictions about residents; instead, their primary concern was to ensure that the committee was following the appropriate procedures when making decisions. Nonetheless, it was still important for these individuals to be experienced and/or to have had some form of leadership training, as part of their role involved navigating potentially challenging power dynamics.

The contribution of program administrators to competency-based medical education often goes unrecognized (Nousiainen et al., 2018). While they are not official members, program administrators usually attend CC meetings to record minutes and/or supplement information where needed. In two of the four CCs, the program administrator had as many speaking turns as the average member. Moreover, observational data revealed that these individuals were often the program director's 'right hand,' filling in gaps when the program director was unable to recall certain details and ensuring that data were for committee members to access and review. Using an information processing lens, one might consider these individuals to be part of the committee's 'memory' system, both in terms of their role in helping to document information for future use and with respect to supplementing important contextual information (Hinsz et al., 1997). The latter function can be compared to the idea of a transactive memory system, which helps increase decision-making accuracy and efficiency through members' facilitation of one another's recollection of events (Mohammed & Dumville, 2001).

5.4.2 Gender and Race/Ethnicity

When the chair and program director were excluded from the analyses (i.e., to account for the effect of position), being a woman or racialized member did not seem to directly impact CC members' contributions. Several factors may account for this, a notable one being that the majority of CC members were clinical faculty who would have already overcome many of the challenges to succeeding in a medical or surgical career and developed positive role expectations (Elsass & Graves, 1997). Regarding gender, it is also possible that women in certain specialties, particularly those that are male dominated,

do not behave according to traditional gender stereotypes (Webster et al., 2016). Although the present study did not test this directly, future studies could answer this question by using a measure of masculinity/femininity rather than gender. Moreover, as colleagues, the majority of CC members would have had prior working relationships with one another, helping to create a collegial atmosphere that was conducive to information sharing. Effective leadership may also have been a factor, as all four CCs had experienced chairs who were skilled at drawing out the contributions of junior members while attempting to curb 'oversharing' by other, more talkative members.

While racialized members were equally as likely to share information as white members, there were some committees on which women contributed more than men. This is particularly interesting when considered in conjunction with the fact that women were overrepresented on three of the four CCs included in the study. Women comprised 40% of the members of CC1, but only 28% of the faculty in the CC's home division. Similarly, women comprised 50% of the members of CC4, but only 33% of faculty. While for CC2 it was not possible to determine the number of women faculty, this number is almost certainly less than 55% due to the underrepresentation of women in surgery (Abelson et al., 2016). These findings suggest that women may be more likely to take on roles related to education than men, which is consistent with the notion of education as 'women's work' (McKinney & Chick, 2010; Myers, 2008). It is noteworthy that in the present study, this phenomenon persisted even in contexts in which women were otherwise underrepresented, such as surgery. The studies in this chapter did not examine whether this was due to women's greater interest in educational roles or training for future

leadership roles, or whether they were disproportionately tasked with such work due to existing stereotypes. Future research may wish to explore this question in further detail.

While there did not appear to be direct effects of gender and race/ethnicity, it is important to explore the possibility of indirect effects by virtue of the committee's leadership. Since the majority of chairs and program directors in this study were white males, and chairs and program directors shared more information than any other member, there may have been an indirect effect of gender and race/ethnicity driven by the information power of members in leadership roles. Since information shared with the committee is inevitably processed through the lens of the individual(s) sharing it, ensuring diversity among individuals in leadership roles (i.e., chairs and program directors) is important for ensuring that a broader range of perspectives are reflected in the committee's decision-making outcomes. It can also positively impact the committee as a whole, as diverse leadership has been linked to greater participation among minority members (Elsass & Graves, 1997; Homan & Greer, 2013). As part of this, appropriate succession planning is necessary to ensure that there are a sufficient number of qualified candidates in the pipeline for future leadership roles.

5.4.3 Bases of Power

The findings of this study suggest that social influence on CCs is driven primarily by information power (French & Raven, 1959; Raven, 1965). Rather than exerting influence through force or reward, certain committee members appeared to have access to more or better information than others, or by virtue of their position, were able to regulate its flow through the committee. Although information power was central to the study

findings, the effects of legitimate and expert power cannot be excluded, since possessing unique information was inextricably linked to being in a position of authority (e.g., chair or program director) or being an external member (e.g., a PhD-trained scientist), although the contributions of these individuals were not specifically explored in this thesis. Moreover, the findings of Chapter 3 suggest that expertise confers the ability to identify which information is predictive of future performance and which is not, which would in turn inform what information individuals such as the program director deemed important to share with their colleagues. Nevertheless, the role of information power appeared most salient in the study findings and is also supported by the interview findings, in which committee members did not pick up on an explicit influence of senior committee members, but commented that the unique "inputs" of these individuals could influence "how the discussion unfolds."

Earlier in this chapter, a potential impact of referent power was also identified; however, the study findings did not suggest that this was particularly salient among the committees studied. Although some members commented on how a shared purpose helped facilitate information sharing within their CC, observational and interview data indicated that members felt comfortable voicing disagreements and were not agreeing for the sake of agreeing. One exception, however, was when committees were faced with time pressures; in these cases, observational data indicated that discussion, including the presentation of information, were sometimes deliberately limited to ensure that the meeting did not impinge on members' time. Future studies should explore whether time

pressures—especially among larger committees with many resident files to review—could lead to greater conformity among members.

5.4.4 Limitations and Future Directions

In the study of speaking turns and number of words spoken, analyses were limited to visual inspection of the data only. In the absence of further statistical analyses, these findings should be interpreted with caution. Moreover, this chapter did not distinguish between primary and backchannel communication, as the latter was more difficult to pick up in audio recordings when CC members were not located near the audio recording device. In future studies, backchannel utterances would be interesting to consider because they may provide additional information about who holds power in a conversation. For example, previous research has shown that women's lesser power in everyday interactions tends to result in a greater likelihood of supporting others (especially men) in conversation through utterances such as, "Mhm," "Yeah," or "Uh-huh" (Dixon & Foster, 1998). Another limitation was that CC members were asked to self-report their gender but not their race/ethnicity. As a result, individuals were classified by the primary investigator, which is less accurate than relying on members' own self-identification.

This chapter also did not explicitly use an intersectional lens; factors that may influence CCs' promotion decisions (i.e., position, gender, and race/ethnicity) were considered separately. Future studies using an intersectional lens would be able to answer questions such as how a member who is junior, a woman, and racialized member might be differentially impacted than a member who, for example, is a white woman with high seniority. Such a qualitative study would also be able to capture individuals' social

identities more holistically (i.e., beyond only their gender, race/ethnicity, or position). Finally, this study only considered the role of anecdotal evidence on the committee's end decisions; however, other types of information shared by members can also influence promotion decisions and should be considered in future studies.

5.5 Conclusion

This chapter explored the role of social hierarchies and power in CC decision-making processes. Analyses of the number of speaking turns and words spoken as well as committees' decisions revealed that members' position had the greatest influence on their contribution to the committee. Specifically, program directors and chairs appeared to have the greatest influence by virtue of their access to information and ability to modulate its flow through the committee, respectively. Gender and race/ethnicity did not appear to have as large of an impact on members' contributions, although gender may have had an indirect role since most chairs and program directors were white men. Interviews with CC members suggested that they had very positive perceptions of the CC decision-making environment and revealed a number of factors conducive to information sharing that may also be applicable to other decision-making bodies.

5.6 References

Abelson, J. S., Chartrand, G., Moo, T.-A., Moore, M., & Yeo, H. (2016). The climb to break the glass ceiling in surgery: Trends in women progressing from medical school to surgical training and academic leadership from 1994 to 2015. *American Journal of Surgery, 212*(4), 566-572. https://doi.org/10.1016/j.amjsurg.2016.06.012

Abramovitch, R., & Grusec, J. E. (1978). Peer imitation in a natural setting. *Child Development, 49*(1), 60-65. https://doi.org/10.2307/1128593

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality & Social Psychology Bulletin, 17*(5), 475-482. https://doi.org/10.1177/0146167291175001

Brooke, M. E., & Ng, S. H. (1986). Language and social influence in small conversational groups. *Journal of Language and Social Psychology, 5*(3), 201-210. https://doi.org/10.1177/0261927X8600500303

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *The University of Chicago Legal Forum*, 139-167.

Delgado, M. R. (2007). Reward-related responses in the human striatum. *Annals of the New York Academy of Sciences, 1104*(1), 70-88. https://doi.org/10.1196/annals.1390.002

Dixon, J. A., & Foster, D. H. (1998). Gender, social context, and backchannel responses. *The Journal of Social Psychology, 138*(1), 134-136. https://doi.org/10.1080/00224549809600364

Elsass, P. M., & Graves, L. M. (1997). Demographic diversity in decision-making groups: The experiences of women and people of color. *The Academy of Management Review, 22*(4), 946-973. https://doi.org/10.2307/259250

Farnan, J. M., Johnson, J. K., Meltzer, D. O., Humphrey, H. J., & Arora, V. M. (2008). Resident uncertainty in clinical decision making and impact on patient care: A

qualitative study. *Quality & Safety in Health Care, 17*(2), 122-126. https://doi.org/10.1136/qshc.2007.023184

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition, 117*(3), 319-331. https://doi.org/10.1016/j.cognition.2010.09.003

French, J. R. P. J., & Raven, B. H. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies in social power* (pp. 150-167). Ann Arbor, MI: Institute for Social Research.

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121*(1), 43-64. https://doi.org/10.1037/0033-2909.121.1.43

Homan, A. C., & Greer, L. (2013). Considering diversity: The positive effects of considerate leadership in diverse teams. *Group Processes & Intergroup Relations, 16*(1), 105-125. https://doi.org/10.1177/1368430212437798

Kelly, J. A., Wildman, H. E., & Urey, J. R. (1982). Gender and sex role differences in group decision-making social interactions: A behavioral analysis. *Journal of Applied Social Psychology, 12*(2), 112-127. https://doi.org/10.1111/j.1559-1816.1982.tb00853.x

Kirchmeyer, C. (1996). Gender roles and decision-making in demographically diverse groups: A case for reviving androgyny. *Sex Roles, 34*(9/10), 649-663. https://doi.org/10.1007/BF01551500

Li, J., Karakowsky, L., & Siegel, J. P. (1999). The effects of proportional representation on intragroup behavior in mixed-race decision-making groups. *Small Group Research, 30*(3), 259-279. https://doi.org/10.1177/104649649903000301

Mays, V. M., & Ghavami, N. (2018). History, aspirations, and transformations of intersectionality: Focusing on gender. In C. B. Travis & J. W. White (Eds.), *APA handbook of the psychology of women: Vol 1. History, theory, and battlegrounds* (pp. 541-566). Washington, DC: American Psychological Association.

McKinney, K., & Chick, N. L. (2010). SoTL as women's work: What do existing data tell us. *International Journal for the Scholarship of Teachiing & Learning 4*(2). https://doi.org/10.20429/ijsotl.2010.040216

Mohammed, S., & Dumville, B. C. (2001). Team mental models in a team knowledge framework: Expanding theory and measurement across disciplinary boundaries. *Journal of Organizational Behavior, 22*(2), 89-106. https://doi.org/10.1002/job.86

Moulton, C.-A., Seemann, N., & Webster, F. (2013). It's all about gender, or is it? *Medical Education, 47*(6), 538-540. https://doi.org/10.1111/medu.12196

Myers, C. B. (2008). College faculty and the scholarship of teaching: Gender differences across four key activities. *Journal of the Scholarship of Teaching and Learning, 8*(2), 38-51.

Ng, S. H., & Bradac, J. J. (1993). *Power in language: Verbal communication and social influence*. Newbury Park, CA: Sage.

Nousiainen, M. T., Mironova, P., Hynes, M., Takahashi, S. G., Reznick, R. K., Kraemer, W., . . . The CBC Planning Committee. (2018). Eight-year outcomes of a

competency-based residency training program in orthopedic surgery. *Medical Teacher, 40*(10), 1042-1054. https://doi.org/10.1080/0142159X.2017.1421751

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1), 1-13. https://doi.org/10.1177/1609406917733847

Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R., & Keltner, D. (2012). Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences of the United States of America, 109*(11), 4086-4091. https://doi.org/10.1073/pnas.1118373109

Ratcliff, N. J., Hugenberg, K., Shriver, E. R., & Bernstein, M. J. (2011). The allure of status: High-status targets are privileged in face processing and memory. *Personality and Social Psychology Bulletin, 37*(8), 1003-1015. https://doi.org/10.1177/0146167211407210

Raven, B. H. (1965). Social influence and power. In M. Fishbein & I. D. Steiner (Eds.), *Current studies in social psychology* (pp. 371-382). New York, NY: Holt, Rinehart, Winston.

Raven, B. H. (1992). A power/interaction model of interpersonal influence: French and Raven thirty years later. *Journal of Social Behavior and Personality, 7*(2), 217-244.

Sheehy-Skeffington, J. (2020). The effects of low socioeconomic status on decision-making processes. *Current Opinion in Psychology, 33*, 183-188. https://doi.org/10.1016/j.copsyc.2019.07.043

Sommers, S. R. (2006). On racial diversity and group decision making: Identifying multiple effects of racial composition on jury deliberations. *Interpersonal Relations and Group Processes, 90*(4), 597-612. https://doi.org/10.1037/0022-3514.90.4.597

Stewart, A. J., & Valian, V. (2018). *An inclusive academy*. Cambridge, MA: MIT Press.

Vanstone, M., & Grierson, L. (2019). Medical student strategies for actively negotiating hierarchy in the clinical environment. *Medical Education, 53*, 1013-1024. https://doi.org/10.1111/medu.13945

Webster, F., Rice, K., Christian, J., Seemann, N., Baxter, N., Moulton, C.-A., & Cil, T. (2016). The erasure of gender in academic surgery: A qualitative study. *American Journal of Surgery, 212*(4), 559-565. https://doi.org/10.1016/j.amjsurg.2016.06.006

Tables and Figures

Table 1. Total number of promotion decisions made by each CC, including the number involving anecdotal evidence and the number in which the outcome appeared to change as a result of the anecdotal evidence provided

Committee	# Promotion Decisions Made	# Involving Anecdotal Evidence	# Changed due to Anecdotal Evidence
CC1	8	4 (50%)	3 (75%)
CC2	54	8 (15%)	1 (13%)
CC3	24	11 (46%)	1 (9%)
CC4	7	6 (86%)	2 (33%)
Total	**93**	**29 (31%)**	**7 (24%)**

Table 2. Position of members providing anecdotes that altered the committee's decision making in each CC

Position	CC1	CC2	CC3	CC4	Total (*n* = 27)
Chair	2	2	0	2	6 (22%)
Program director	2	6	1	0	9 (33%)
Other senior member	0	0	3	0	3 (11%)
Regular member	0	3	0	4	7 (26%)
Junior member	0	0	1	0	1 (4%)
Academic advisor	1	0	0	0	1 (4%)

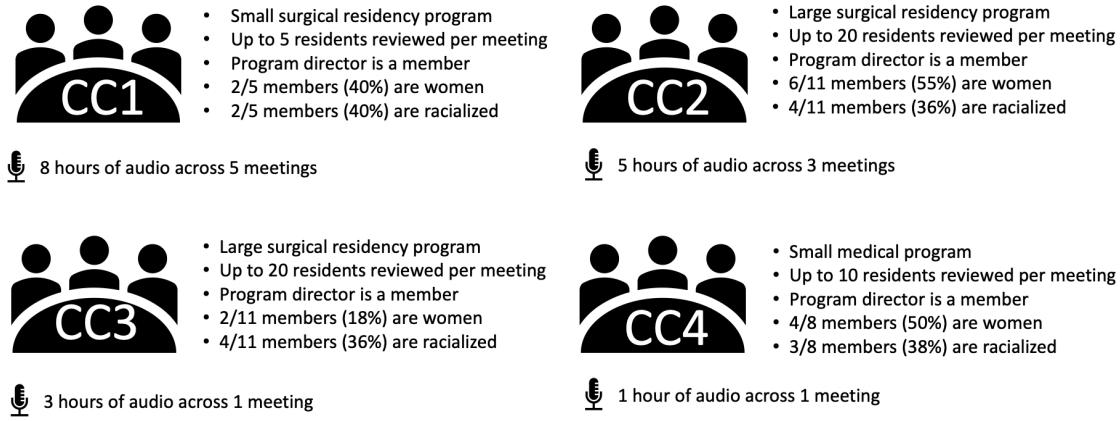

Figure 1. Characteristics of the four CCs from which meeting transcripts were obtained, along with the amount of data available for each

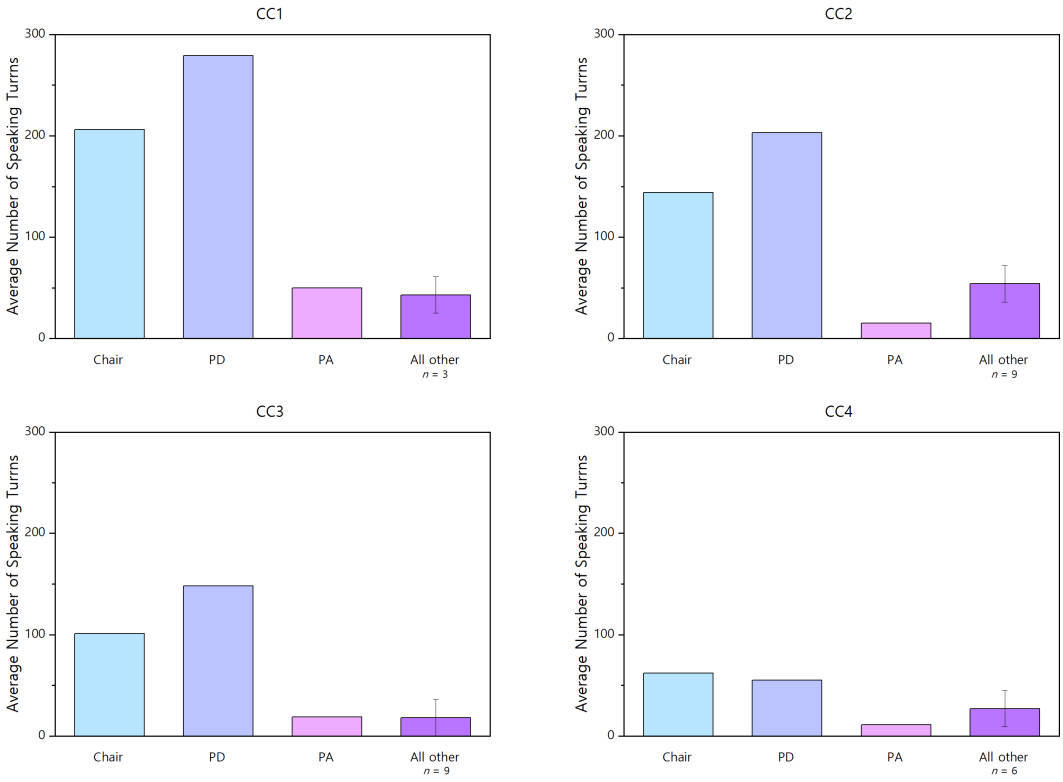

Figure 2. Average number of speaking turns per meeting across the four CCs included in the study, stratified by members' position on the committee. Error bars represent standard errors of the mean. PD = Program director, PA = Program administrator, All other = all other members

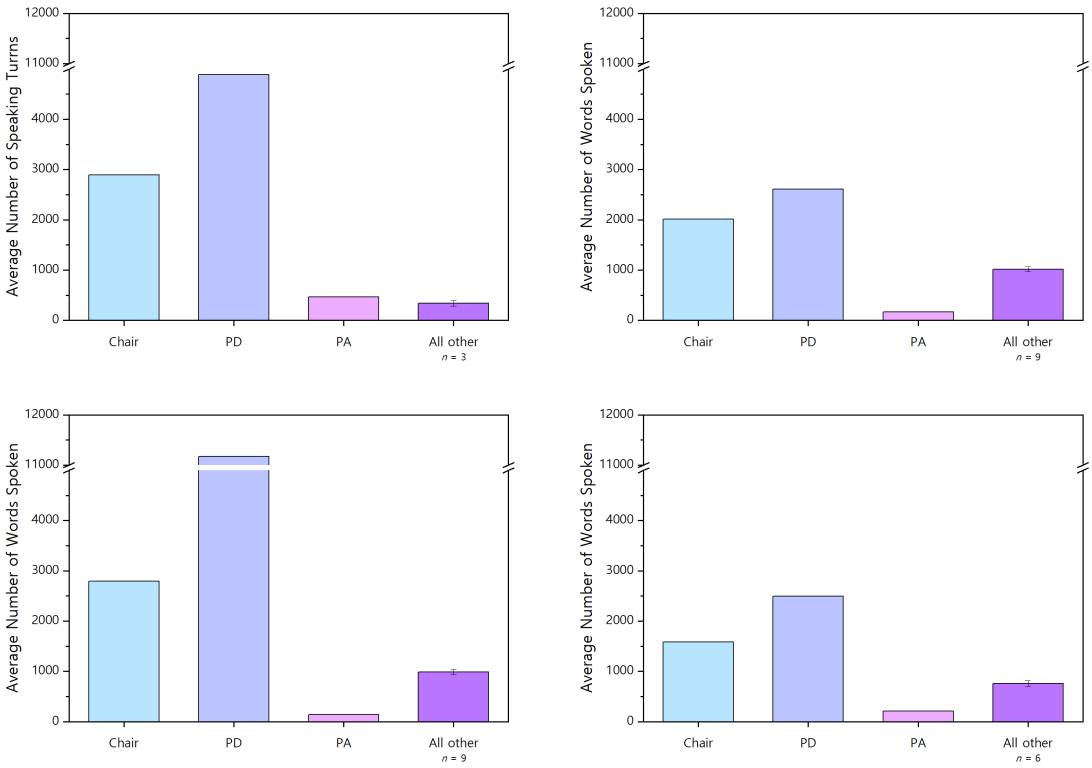

Figure 3. Average number of words spoken per meeting across the four CCs included in the study, stratified by members' position on the committee. Error bars represent standard errors of the mean. PD = Program director, PA = Program administrator, All other = all other members

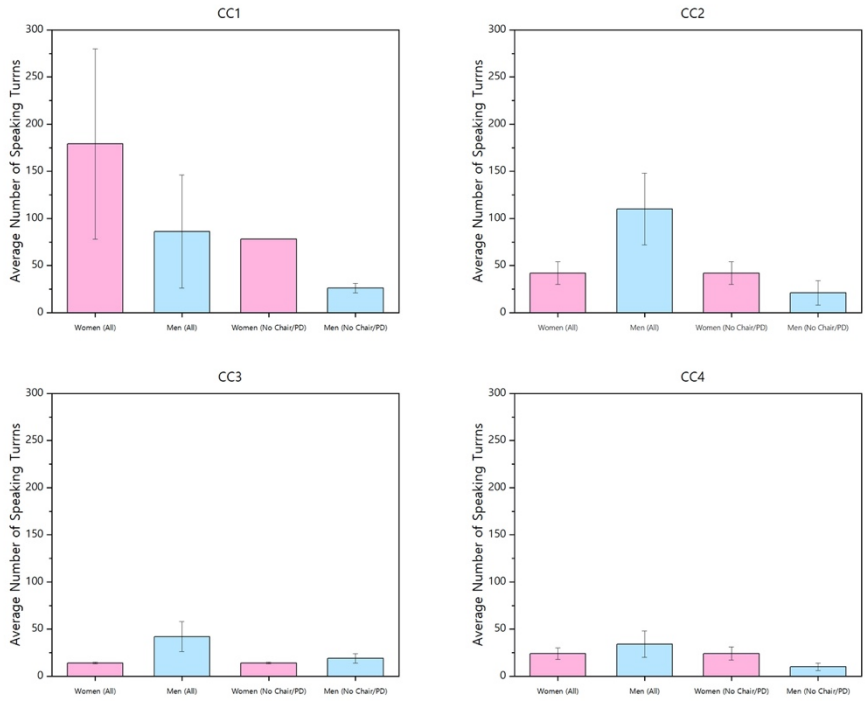

Figure 4. Average number of speaking turns per meeting across the four CCs included in the study, stratified by members' gender. Error bars represent standard errors of the mean. The first set of bars represents all members while the second set represents membership with the chair and program director removed

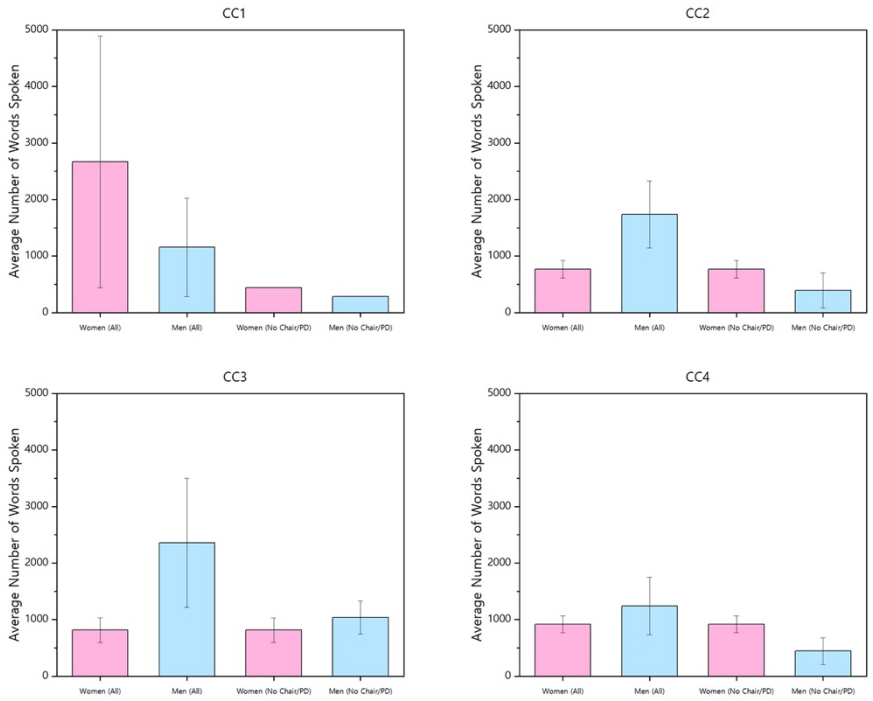

Figure 5. Average number of words spoken per meeting across the four CCs included in the study, stratified by members' gender. Error bars represent standard errors of the mean. The first set of bars represents all members while the second set represents membership with the chair and program director removed

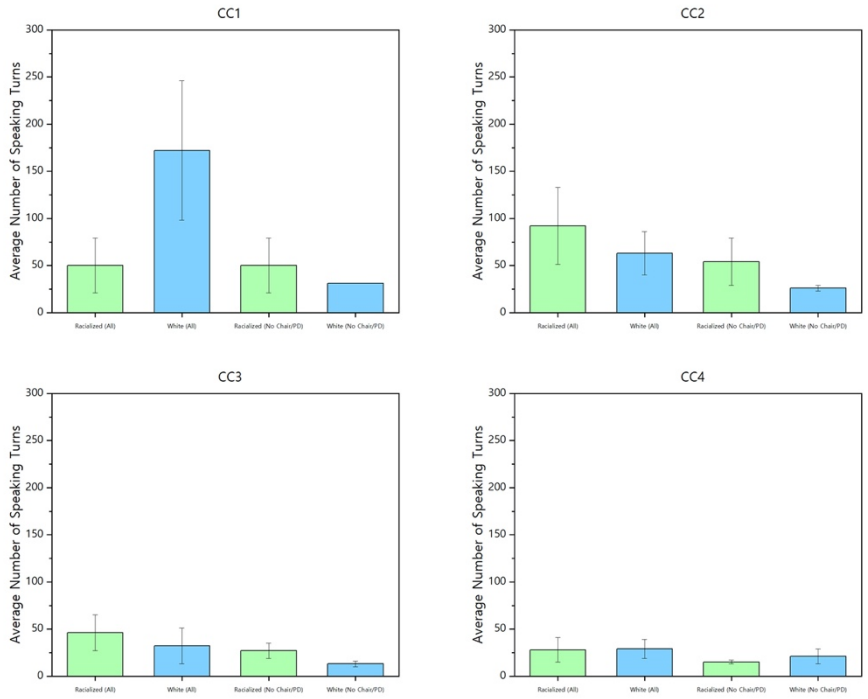

Figure 6. Average number of speaking turns per meeting across the four CCs included in the study, stratified by members' race/ethnicity. Error bars represent standard errors of the mean. The first set of bars represents all members while the second set represents membership with the chair and program director removed

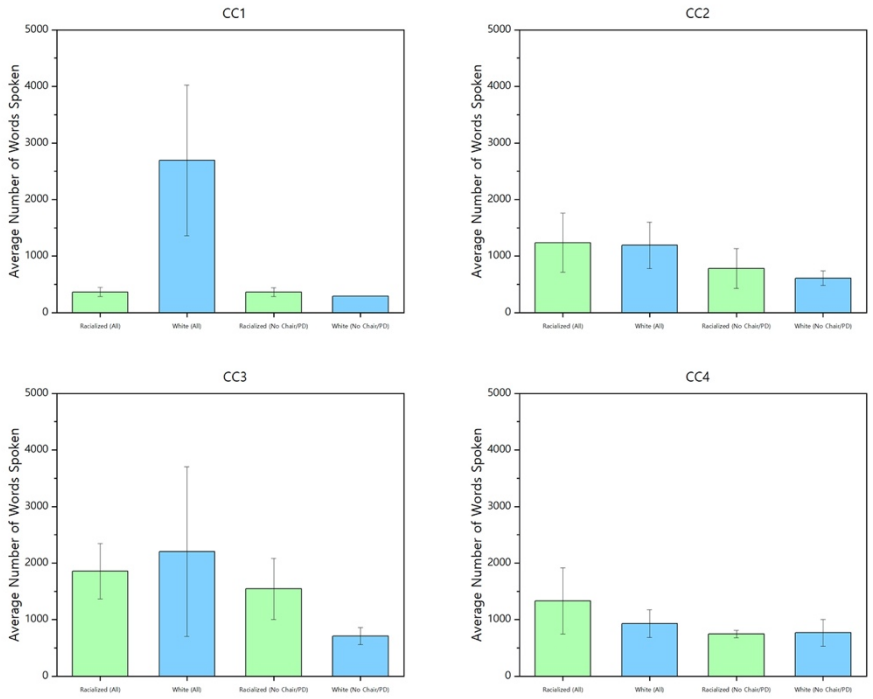

Figure 7. Average number of words spoken per meeting across the four CCs included in the study, stratified by members' race/ethnicity. Error bars represent standard errors of the mean. The first set of bars represents all members while the second set represents membership with the chair and program director removed

Chapter 6

General Discussion

6.1 Overview of Thesis and Key Findings

This thesis comprised two main sections. The first section (Chapter 2) explored CC implementation practices at a Canadian institution while the second (Chapters 3 to 5) delved into their decision-making processes—specifically, how they process anecdotal evidence. Chapter 2 was a longitudinal study consisting of interviews, observations, and surveys to examine CC practices over a three-year period. While participants' overall perceptions of CCs were positive, this study revealed a number of implementation challenges to be addressed going forward, including optimizing membership, maintaining capacity among members, engaging faculty and residents in the CBD process, sharing and aggregating data, and developing a clear mandate. The findings also suggested that stakeholders had a limited understanding of how CCs make decisions.

Chapter 3 described how novice and experienced individual raters make competence judgments involving anecdotal evidence. It also explored novice raters' reasoning for their promotion decisions. While both novice and experienced raters incorporated anecdotal evidence into their decision making, novice raters were more sensitive to this information than experienced raters. Both groups displayed variability in their promotion decisions, particularly when anecdotal evidence was involved; however, this was once again more pronounced among novice raters than experienced raters. A final experiment on reasoning revealed that participants' decision to incorporate anecdotal evidence into their decision making was generally deliberate and within their conscious awareness.

Chapter 4 explored raters' decision-making processes in groups. Novice and experienced raters completed the same experiment as in Chapter 3 in groups of approximately five. In both cases, the group decision-making process helped remove some of the unwanted variability in individual raters' decisions. However, groups remained sensitive to anecdotal evidence under certain conditions (e.g., unprofessional behaviour), suggesting that they may be able to discern when this information could be conveying something important and should not be ignored. The value of anecdotal evidence was further elucidated in an analysis of CC meeting and interview transcripts, which revealed a number of direct and indirect roles of anecdotal evidence including corroborating, explaining, questioning, substituting, and prompting additional action.

Chapter 5 delved further into the social processes underlying CCs' group decisions by exploring the roles of social hierarchies and power. Analyses of the number of speaking turns and words spoken revealed that members' position on the committee had a much greater influence on their contributions to the committee than other social categories, such as members' gender or race/ethnicity. The chair and program director were found to have the most influence on the committee in terms of the number of speaking turns and words spoken. These results were corroborated by an analysis of decisions made by CCs, which showed that anecdotal evidence provided by program directors had the greatest influence on committees' decisions. Social influence was found to be guided mostly by informational power, with the chair serving as a modulator of information flow throughout the committee and the program director having access to unique information not possessed by other members. Despite the disproportionate influence of certain

members on CC decision-making processes, junior members had very positive perceptions of the power dynamics on CCs and felt as if they could freely share their thoughts and opinions. This may be because information power is a more subtle form of influence that does not necessarily diminish the contributions of other members. Interviews with CC members revealed a number of characteristics that helped facilitate robust information sharing in CCs: careful selection of members, a prior history of working together, effective leadership, a shared purpose, and specific decision-making procedures such as review assignments and data-driven discussion.

6.2 The Role of Groups in Solving Complex Problems

Although the research in this thesis was done in the context of CCs, the findings have implications for group decision making more broadly. Complexity is unavoidable in most human decision-making tasks, including in medicine (Saunders, 2000). As a result, there can be substantial variability in the outcomes of decisions made by individuals, as observed in Chapter 3. This lack of consistency can make decisions seem less defensible to stakeholders, and therefore more likely to be contested and/or not taken seriously (Lunenburg, 2010; van der Bles, van der Linden, Freeman, & Spiegelhalter, 2020).

Groups offer a potential solution for dealing with complex decision-making tasks. Indeed, their prevalence among both humans and animals suggests that groups can offer a robust way of reconciling multiple perspectives to arrive at the best decision alternative (Kameda, Wisdom, Toyokawa, & Inukai, 2012). Involving multiple experts in a decision-making task means that there is not only more information (i.e., knowledge) that can be shared, but also more lenses through which this information can be scrutinized (Hinsz,

Tindale, & Vollrath, 1997; Mesmer-Magnus & DeChurch, 2009; Stasser, 1999). The benefits of groups observed in this thesis have the potential to apply to a range of contexts beyond CCs. As noted earlier, there is a growing recognition of the importance of groups in medicine, which includes not only CCs, but a range of other groups and teams involved in diagnosing and treatment complex conditions, such as cancer (Wright, De Vito, Langer, Hunter, & The Expert Panel on Multidisciplinary Cancer Conference Standards, 2007), or performing complex procedures, such as surgeries (Sebok-Syer et al., 2018).

As relatively well-functioning groups, CCs offer a model that can be used to promote positive group interactions. For example, Chapter 5 revealed several factors that helped CCs enact their mandate, including careful selection of members, prior history of working together, effective leadership, a shared purpose, and specific decision-making procedures. Many of these ideas are broadly applicable to other groups in medicine. For example, groups comprising members who do not know each other very well should provide members with an opportunity to get to know one another and build proficiency in working together, perhaps through a training session involving a simulated decision-making task (Gilley, Morris, Waite, Coates, & Veliquette, 2010).

Effective leadership is also important for facilitating positive group interactions (Ezziane et al., 2012; Gilley et al., 2010). Traits that were observed to be particularly effective among CCs was having excellent listening and facilitation skills; having a strong understanding of the committee's mandate and decision-making procedures; and being able to navigate the power dynamics inherent within the committee, which often required a certain amount of leadership experience. Another factor that helped promote effective

information sharing within CCs was distributing tasks among members. For example, tasking each committee member with reviewing a certain number of resident files had a positive effect on promoting member participation in the group decision-making process and also helped ensure that the workload was more evenly distributed among members.

Linking all conversation to specific data points (i.e., evidence in a resident's file) also helped some members feel more comfortable sharing their opinions. Some CC members noted that data-driven decision making helped them felt less connected to their relationship with the resident and more to what was being evaluated (i.e., their performance). This is consistent with guidelines from regulatory bodies such as the Royal College of Physicians and Surgeons of Canada, which emphasize the importance of data-driven discussion in CC decision-making processes (Royal College of Physicians and Surgeons of Canada, 2018). It is possible that as CCs become better integrated into residency training programs, stakeholders will begin to recognize that it is data-drivenness, rather than objectivity (per Chapter 2), that results in better decision making.

6.3 A Rationale for Equity, Diversity, and Inclusion

While issues related to equity, diversity, and inclusion (EDI) are longstanding in medicine, the current sociopolitical climate has reemphasized their importance (Canadian Medical Association, 2019; Yancy, 2020). The findings of this thesis provide a rationale for EDI that moves beyond simply a moral imperative (i.e., because it is the right thing to do), and towards an understanding that a more diverse workforce also means better decision making. Aside from better decision making, greater diversity is also aligned with

societal expectations of the medical profession, as indicated in multiple sections of the CanMEDS physician competency framework (Frank, Snell, & Sherbino, 2015).

Chapter 4 showed that group responses were heavily dependent on members' initial preferences. The latter finding is well-supported in the group decision-making literature (Kerr, MacCoun, & Kramer, 1996; Stasser, 1999). An obvious implication of this finding is that a group's composition will inform the decisions that it makes. This speaks to the importance of optimizing CC membership, a theme previously identified in Chapter 2. If a committee is comprised solely of members with similar backgrounds, then they will likely think very similarly and produce a decision outcome that is reflective of their views but not necessarily those of whom they serve (Janis, 1982). Increasing the diversity of committees to include members from a variety of backgrounds (i.e., both demographic and functional specialization) means that there is more unique information to be shared between members, and more lenses through which to scrutinize this information (Hinsz et al., 1997; Mesmer-Magnus & DeChurch, 2009; Stasser, 1999). Thus, the end decisions made by the committee will likely be of higher-quality and reflect a broader range of perspectives. This is important for ensuring the defensibility of decisions, particularly in situations that are not captured by pre-existing decision-making criteria.

However, diversity alone is not enough. As previously mentioned, a number of factors must be in place to ensure that members feel included in the group decision-making process. A number of these factors were uncovered in Chapter 5, including those covered in 6.2 of the present chapter. Another important factor is how diversity is managed within the group. Including a single 'token' individual to represent a given social group is

unlikely to facilitate positive change, as this individual would likely feel isolated and uncomfortable speaking up (Elsass & Graves, 1997). Ensuring that diversity is represented in various forms on the committee (e.g., gender, seniority, race/ethnicity, position, etc.) is more conducive to success (Page, 2007). Group composition should also take into account intersectionality, the notion that social categories such as race, class, and gender are highly interconnected and interdependent (Crenshaw, 1989).

The group leader also has an important role in facilitating an inclusive decision-making process (Homan & Greer, 2013). First and foremost, group leaders are responsible for ensuring that CC members are appropriately trained to perform their roles. To this end, CC chairs may consider having their members complete unconscious bias training as part of the onboarding process and facilitate reflexivity throughout the committee's deliberations. Furthermore, as observed in Chapter 5, group leaders have an important responsibility to ensure that all members feel comfortable contributing to the conversation and that more talkative members do not overpower those who are quieter. Finally, while group leaders have a direct influence on the group decision-making process, they can also affect efforts indirectly: members of teams with diverse leadership may feel more included—and therefore more willing to contribute—if they see themselves represented among the group's upper ranks (Elsass & Graves, 1997).

6.4 Do Competence Committees Justify the Increased Cost?

After exploring broader implications, this chapter will now turn to some of the more specific implications of this thesis. While the potential for CCs to improve decision-making processes related to resident promotion have been extensively highlighted in the

literature, the majority of these studies have been theoretical and/or based on findings from a single programmatic context. This thesis makes a significant contribution to the literature as it is the first, to the primary investigator's knowledge, to study CCs from across multiple programs whilst also triangulating data from multiple sources.

The findings of this thesis confirm that the CCs studied as part of this research are, in fact, a positive addition to residency training programs. This is important because CCs require a substantial time commitment from clinicians whose primary duty is to provide patient care; thus, ensuring that they confer decision-making benefits is critical for justifying their use as part of CBME curricula. While the benefit of CCs is clear, the findings of this thesis revealed potential misconceptions about *why* they are beneficial. For example, participants in Chapter 2 often commented on the fact that CCs were superior to prior promotion and review processes because of their reliance on more 'objective' data sources, such as Entrustable Professional Activities (EPAs). The findings of this thesis suggest that this assumption is misguided, as CCs rely on a variety of data sources when making decisions that include but are certainly not limited to EPAs. Indeed, the use of a variety of data sources to make decisions is recognized as part of their mandate (Royal College of Physicians and Surgeons of Canada, n.d.-a).

In certain situations, anecdotal evidence, which is clearly subjective, can be as, if not more, useful than EPAs for contextualizing assessment data and understanding a resident's performance. The true benefit of CCs lies in the fact that they not only have access to more information (i.e., by virtue of having multiple members), but they can also more thoroughly scrutinize this information and decide how much emphasis to place on it

through discussion among multiple experts. This aligns with existing accounts of the CC decision-making process that position data interpretation and problem solving at the core of CCs' work (Pack, Lingard, Watling, Chahine, & Cristancho, 2019; Pack, Lingard, Watling, & Cristancho, 2020). It is also consistent with other literature on group decision making, which emphasize the value of groups in solving complex problems (Kerr & Tindale, 2004; Laughlin, Hatch, Silver, & Boh, 2006).

6.5 The Role of Anecdotal Evidence in Competence Committee Decision-Making Processes

Given the strong emphasis of the existing literature on more conventional data sources, such as EPAs, this thesis considered the role that anecdotal evidence plays in CC decision-making processes. Findings from Chapters 3 and 4 revealed that anecdotal evidence plays an important role in CC decision-making processes. While it could occasionally stand in for other, missing assessment data, anecdotal evidence was most often responsible for modulating how other data were interpreted by providing context that enabled committee to make more robust decisions about resident performance.

These findings suggest a need to explicitly recognize anecdotal evidence as a legitimate data source for CCs, as well as to devise appropriate procedures for how this information is to be handled (e.g., in CC guidelines or terms of reference). While anecdotal evidence may motivate a search for additional evidence or points of triangulation, it—like any other singular data point—should not be used to make promotion decisions on its own. If an anecdote is shared, committees should seek to ensure that other members have the opportunity to contribute their perspectives, a strategy often used by experienced chairs of the CCs observed in this thesis. The source of any

anecdotes should also be thoroughly scrutinized by the committee—e.g., by asking, "Is this a trusted individual? What perspectives might inform their observations and interpretations of the resident's performance?" The committee may also wish to search for additional, documented evidence in a resident's file that might corroborate the anecdote(s) being shared or to speak to the resident directly to ascertain their perspective on their performance and how best to facilitate their progress moving forward. Once a decision is made, CCs should ensure that the committee's rationale, as well as any evidence used to come to the decision, are clearly documented in the meeting minutes (or in a separate document) in case it is necessary to revisit the decision in the future.

6.6 Next Steps

Although the findings of this thesis revealed generally positive perceptions of CCs, it also identified a number of outstanding implementation challenges that were pervasive across multiple programs. These include optimizing membership, maintaining capacity among members, engaging faculty and residents in the CBD process, sharing and aggregating data, and developing a clear mandate. As previously mentioned, a strength of this thesis was that multiple programs were studied over time to develop an understanding of CC implementation practices. It is recommended that future studies also take a similar approach, as policy decisions are often made at the institutional and/or national levels, meaning that they affect multiple programs (Andolsek, Padmore, Hauer, Edgar, & Holmboe, 2020; Royal College of Physicians and Surgeons of Canada, n.d.-b). Understanding what challenges are pervasive across multiple programs is also important because some implementation challenges, such as the paucity of robust data-sharing

platforms, would also benefit from a more concerted approach to avoid duplicating efforts that may require a significant amount of time and resources.

This thesis also highlighted the importance of using multiple methods to study complex phenomena, such as CCs. The experimental work in this thesis allowed for specific variables to be manipulated in order to specifically isolate the effect of anecdotal evidence. However, these data were much more meaningful when considered in conjunction with observational data and findings from interviews and surveys conducted with CC members. Future work would also benefit from a similar, multi-method approach to triangulate evidence from different sources in order to build a more robust understanding of group decision-making processes in medicine.

In the future, there remains ample room to continue expanding the medical education community's understanding of CCs. From an implementation standpoint, it will be important to consider how CCs interface with various data-sharing platforms as they become available, and how these tools can be used to make decision-making processes more efficient. Attention should also be directed to capacity building (including across programs) and succession planning, to ensure that CCs have an appropriate set of members to draw upon as the current membership begins to turn over. Certain aspects of CCs' mandate remain open to interpretation, including the extent to which they should be involved in the development and oversight of educational and remediation plans for trainees, areas that should be clarified as CCs continue to mature.

With respect to CC decision-making processes, future research might focus on more complex portfolios, with multiple sources of data, in order to understand how this

information is interpreted by CC members. This would also provide insight into how different data sources are weighted by CC members under different conditions, as well as whether certain data sources are regarded as more credible than others. The program director role is also worth following over time to determine whether their influence remains as high as in this thesis or whether more robust data-sharing mechanisms reduce the reliance on their perspectives. Finally, one might consider how the decision-making processes of CCs compare to other decision-making processes in medicine, such as those of multidisciplinary cancer conferences or trauma/surgical teams.

6.7 References

Andolsek, K., Padmore, J., Hauer, K. E., Edgar, L., & Holmboe, E. (2020). Clinical competency committees: A guidebook for programs (3rd edition). Retrieved from https://www.acgme.org/Portals/0/ACGMEClinicalCompetencyCommitteeGuidebook.pdf

Canadian Medical Association. (2019). CMA policy: Equity and diversity in medicine. Retrieved from https://policybase.cma.ca/documents/Policypdf/PD20-02.pdf

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *The University of Chicago Legal Forum*, 139-167.

Elsass, P. M., & Graves, L. M. (1997). Demographic diversity in decision-making groups: The experiences of women and people of color. *The Academy of Management Review, 22*(4), 946-973. https://doi.org/10.2307/259250

Ezziane, Z., Maruthappu, M., Gawn, L., Thompson, E. A., Athanasiou, T., & Warren, O. J. (2012). Building effective clinical teams in healthcare. *Journal of Health Organization and Management, 26*(4), 428-436. https://doi.org/10.1108/14777261211251508

Frank, J. R., Snell, L., & Sherbino, J. (2015). *CanMEDS 2015 physician competency framework*. Retrieved from http://www.royalcollege.ca/rcsite/documents/canmeds/canmeds-full-framework-e.pdf

Gilley, J. W., Morris, M. L., Waite, A. M., Coates, T., & Veliquette, A. (2010). Integrated theoretical model for building effective teams. *Advances in Developing Human Resources, 12*(1), 7-28. https://doi.org/10.1177/1523422310365309

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121*(1), 43-64. https://doi.org/10.1037/0033-2909.121.1.43

Homan, A. C., & Greer, L. (2013). Considering diversity: The positive effects of considerate leadership in diverse teams. *Group Processes & Intergroup Relations, 16*(1), 105-125. https://doi.org/10.1177/1368430212437798

Janis, I. L. (1982). *Groupthink (2nd ed.)* Boston, MA: Houghton Mifflin.

Kameda, T., Wisdom, T., Toyokawa, W., & Inukai, K. (2012). Is consensus-seeking unique to humans? A selective review of animal group decision-making and its implications for (human) social psychology. *Group Processes & Intergroup Relations, 15*(5), 673-689. https://doi.org/10.1177/1368430212451863

Kerr, N. L., MacCoun, R. J., & Kramer, G. P. (1996). Bias in judgment: Comparing individuals and groups. *Psychological Review, 103*(4), 687-719. https://doi.org/10.1037/0033-295X.103.4.687

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review of Psychology, 55*, 623-655. https://doi.org/10.1146/annurev.psych.55.090902.142009

Laughlin, P. R., Hatch, E. C., Silver, J. S., & Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: Effects of group size. *Journal of Personality and Social Psychology, 90*(4), 644-651. https://doi.org/10.1037/0022-3514.90.4.644

Lunenburg, F. C. (2010). Group decision making. *National Forum of Teacher Education Journal, 20*(3).

Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology, 94*(2), 535-546. https://doi.org/10.1037/a0013773

Pack, R., Lingard, L., Watling, C. J., Chahine, S., & Cristancho, S. M. (2019). Some assembly required: Tracing the interpretive work of clinical competency committees. *Medical Education, 53*(7), 723-734. https://doi.org/10.1111/medu.13884

Pack, R., Lingard, L., Watling, C. J., & Cristancho, S. (2020). Beyond summative decision-making: Illuminating the broader roles of competence committees. *Medical Education, 54*(6), 517-527. https://doi.org/10.1111/MEDU.14072

Page, S. E. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and society*. Princeton, NJ: Princeton University Press.

Royal College of Physicians and Surgeons of Canada. (2018). Competence committee guideline: Process and procedures in decision making. Retrieved from http://www.royalcollege.ca/rcsite/documents/cbd/competence-committees-process-procedures-e

Royal College of Physicians and Surgeons of Canada. (n.d.-a). Competence committees. Retrieved from http://www.royalcollege.ca/rcsite/cbd/assessment/competence-committees-e

Royal College of Physicians and Surgeons of Canada. (n.d.-b). Competence committees—Guidelines for the terms of reference. Retrieved from http://www.royalcollege.ca/rcsite/cbd/assessment/committees/managing-competence-committees-e

Saunders, J. (2000). The practice of clinical medicine as an art and as a science. *Medical Humanities, 26*(1), 18-22. https://doi.org/10.1136/mh.26.1.18

Sebok-Syer, S. S., Chahine, S., Watling, C. J., Goldszmidt, M., Cristancho, S., & Lingard, L. (2018). Considering the interdependence of clinical performance: Implications for assessment and entrustment. *Medical Education, 52*(9), 970-980. https://doi.org/10.1111/medu.13588

Stasser, G. (1999). A primer of social decision scheme theory: Models of group influence, competitive model-testing, and prospective modeling. *Organizational Behavior*

and Human Decision Processes, 80(1), 3-20. https://doi.org/10.1006/obhd.1999.2851

van der Bles, A. M., van der Linden, S., Freeman, A. L. J., & Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences of the United States of America, 117*(14), 7672-7683. https://doi.org/ https://doi.org/10.1073/pnas.1913678117

Wright, F. C., De Vito, C., Langer, B., Hunter, A., & The Expert Panel on Multidisciplinary Cancer Conference Standards. (2007). Multidisciplinary cancer conferences: A systematic reviewand development of practice standards. *European Journal of Cancer, 43*(6), 1002-1010. https://doi.org/10.1016/j.ejca.2007.01.025

Yancy, C. W. (2020). Academic medicine and Black Lives Matter: Time for deep listening. *Journal of the American Medical Association*. https://doi.org/10.1001/jama.2020.12532

Appendix 1 – Phase 1 Survey (Pre-Implementation)

1. I am aware of what competence committees are and their intended role in the new Competency by Design curriculum.

[ ] Strongly agree
[ ] Agree
[ ] Neither agree nor disagree
[ ] Disagree
[ ] Strongly disagree

2. Competence committees will improve educational outcomes for residency training programs.

[ ] Strongly agree
[ ] Agree
[ ] Neither agree nor disagree
[ ] Disagree
[ ] Strongly disagree

3. Competence committees will improve decision making processes around resident promotion.

[ ] Strongly agree
[ ] Agree
[ ] Neither agree nor disagree
[ ] Disagree
[ ] Strongly disagree

4. Faculty/staff outside a given residency program (i.e., external members) should be included as competence committee members.

[ ] Strongly agree
[ ] Agree
[ ] Neither agree nor disagree
[ ] Disagree
[ ] Strongly disagree

5. Laypeople (i.e., patients and/or community members) should be included as competence committee members.

[ ] Strongly agree
[ ] Agree
[ ] Neither agree nor disagree

[ ] Disagree
[ ] Strongly disagree

6. Residents should be included as competence committee members.

[ ] Strongly agree
[ ] Agree
[ ] Neither agree nor disagree
[ ] Disagree
[ ] Strongly disagree

7. Competence committees will require more commitment from faculty/staff than current promotion processes.

[ ] Strongly agree
[ ] Agree
[ ] Neither agree nor disagree
[ ] Disagree
[ ] Strongly disagree

8. My program has the appropriate tools and resources to implement competence committees.

[ ] Strongly agree
[ ] Agree
[ ] Neither agree nor disagree
[ ] Disagree
[ ] Strongly disagree
[ ] I don’t know

9. My program has taken steps towards implementing competence committees.

[ ] Strongly agree
[ ] Agree
[ ] Neither agree nor disagree
[ ] Disagree
[ ] Strongly disagree
[ ] I don’t know

10. Setting up a competence committee in my program seems overwhelming.

[ ] Strongly agree
[ ] Agree
[ ] Neither agree nor disagree

[ ] Disagree
[ ] Strongly disagree
[ ] Not applicable; I am not responsible for setting up the committee

11. Overall, I feel that competence committees are a good idea.

[ ] Strongly agree
[ ] Agree
[ ] Neither agree nor disagree
[ ] Disagree
[ ] Strongly disagree

12. Are you currently a member of a competence committee?

[ ] Yes
[ ] No

13. [If yes to Q12] I feel that I have the appropriate tools and resources to carry out my role as a competence committee member. (If you are a part of more than one competence committee, please think of the committee on which you have the most experience.)

[ ] Strongly agree
[ ] Agree
[ ] Neither agree nor disagree
[ ] Disagree
[ ] Strongly disagree

14. [If yes to Q12] I feel that the committee will be effective in working together to make decisions about resident promotion. (If you are a part of more than one competence committee, please think of the committee on which you have the most experience.)

[ ] Strongly agree
[ ] Agree
[ ] Neither agree nor disagree
[ ] Disagree
[ ] Strongly disagree

Appendix 2 – Phase 1 Interview Guide (Pre-Implementation)

1. Can you give us a little background about yourself—e.g., speciality, current position, location, etc.?

2. To the best of your knowledge, how does your program currently make decisions with respect to promoting residents to the next level of training? How regularly is each resident reviewed throughout their training and by whom? [*Follow-up*] What do you see as the advantages and disadvantages of the current system that's used? How consistent do you think the process is across trainees?

3. Can you tell us what you know about competence committees? [*Prompt*] For example, what do you see as their intended purpose? How will they be structured in terms of membership, reporting structure, etc.?

4. Based on what you currently know about them, what are your thoughts on/attitudes towards competence committees? What do you see as their advantages and disadvantages? [*Follow-up*] What seems to be the general perception towards competence committees among your colleagues?

5. What, if any, qualifications should competence committee members have? [*Follow-up*] What, if any, specific training should they receive?

6. Do you see a role for laypeople (i.e., community members, patients, etc.) on competence committees? If so, what would that role be? Do you perceive any specific advantages or disadvantages of including these individuals as members? [*Follow-up*] What, if any, specific training should they receive?

7. Should resident input be factored into decisions made by competence committees? If so, how? [*Note: It is not currently recommended that residents be members*]

8. Has your program taken any steps towards implementing competence committees? If so, what have these steps entailed?

9. If your program has taken steps towards implementing competence committees, what has there been anything you feel that's worked particularly well? Have you encountered any challenges or roadblocks? [*Follow-up*] What resources do you think would be helpful for addressing some of these challenges/roadblocks?

10. If your program has not yet taken any steps towards implementing competence committees, do you anticipate any challenges or roadblocks? [*Follow-up*] What resources do you think would be helpful for addressing some of these challenges/roadblocks?

11. Is there anything else that you'd like to tell us about competence committees that you feel would be useful as they are implemented across residency training programs at McMaster?

Appendix 3 – Phase 2 Survey (One Year Mark)

Administrative

1. What is the name of your program?

2. How many residents are in your program?

[ ] 1-4
[ ] 5-9
[ ] 10-10
[ ] 20-29
[ ] 30-39
[ ] 40+

3. When did or will your program transition to Competence by Design?

[ ] July 2017
[ ] July 2018
[ ] July 2019
[ ] July 2020
[ ] July 2021
[ ] July 2022

4. Do you have a Competence Committee?

[ ] Yes
[ ] No (Please comment)

5. How many members on your Competence Committee?

[ ] 1-4
[ ] 5-9
[ ] 10+

6. When was your Competence Committee initiated?

7. Are the members of your committee:

[ ] Representative of a particular constituency of the program
[ ] Chosen for their experience in education
[ ] Both of the above
[ ] None of the above (Please describe)

8. Is there a resident representative on the Competence Committee?

[ ] Yes (How are they selected?)
[ ] No (How was this decision made?)

9. Is there a non-program representative who is "external" to the teaching faculty on your Competence Committee?

[ ] Yes (Please elaborate if you wish)
[ ] No (Please elaborate if you wish)

10. If yes, please check the appropriate description of that/those individual(s):

[ ] Program director from another residency program at McMaster
[ ] Faculty member from another residency program at McMaster
[ ] Allied health professional
[ ] Researcher
[ ] Public member
[ ] Other

11. How often does your Competence Committee meet?

[ ] Quarterly
[ ] Monthly
[ ] Bi-monthly
[ ] Ad hoc
[ ] Other (Please describe)

12. Do you have an orientation for new Competence Committee members?

[ ] Yes (Please describe)
[ ] No (Please elaborate if you wish)

13. Are committee members permitted to attend meetings by teleconference?

[ ] Yes (Please elaborate if you wish)
[ ] No (Please elaborate if you wish)

14. Does your program have academic coaches?

15. Describe how your academic coaches are selected.

16. Are your academic coaches members of the Competence Committee?

[ ] Yes (Please elaborate if you wish)
[ ] No (Please elaborate if you wish)

Information Sharing / Technology

17. How are you sharing resident assessment data?

[ ] Electronically
[ ] Paper
[ ] Both of the above

18. If electronic, what platform(s) are you using? (Select all that apply)

[ ] MedSIS
[ ] RCPSC Resident E-Portfolio
[ ] Medportal
[ ] Locally developed platform
[ ] Other
[ ] N/A

File Review

19. Is there a set time for file review?

[ ] Yes (Please elaborate if you wish)
[ ] No (Please elaborate if you wish)

20. If yes, how much time is allocated per review?

[ ] 0-2 minutes
[ ] 3-5 minutes
[ ] 6-10 minutes
[ ] More than 10 minutes

Please elaborate if you wish.

21. Does the Chair review every file?

[ ] Yes (Please elaborate if you wish)
[ ] No (Please elaborate if you wish)

22. Does the Chair act as the primary or secondary reviewer for any files?

[ ] Yes (Please elaborate if you wish)
[ ] No (Please elaborate if you wish)

23. Does the Chair review only files of residents not meeting EPAs/milestones?

[ ] Yes (Please elaborate if you wish)
[ ] No (Please elaborate if you wish)

24. Does the Chair **NOT** review any files at all?

[ ] Yes (Please elaborate if you wish)
[ ] No (Please elaborate if you wish)

25. How is it decided which residents will be reviewed at each meeting? Please describe.

26. How are the files of the residents who are meeting their EPAs and milestones discussed?

[ ] They are discussed at meetings and given more time than residents not meeting their EPAs and milestones
[ ] They are discussed at meetings and given the same amount of time as residents not meeting their EPAs and milestones
[ ] They are discussed at meetings and given less time than residents not meeting their EPAs and milestones
[ ] They are not discussed at meetings but reviewed prior to the meeting and agreed to by consensus
[ ] Other (Please describe)

27. Is there a rationale for the order in which residents' files are reviewed? Please describe.

28. How many files are assigned to committee members?

[ ] Each committee members reviews all files
[ ] Each committee member is assigned files to review (Specify number)
[ ] Other (Please describe)

29. Are committee members permitted to review the same resident at subsequent meetings?

[ ] Yes (Please elaborate if you wish)
[ ] No (Please elaborate if you wish)

Discussion of the Committee

30. When are members able to raise questions at the presentation?

[ ] Throughout the reviewer's presentation
[ ] At the end of the reviewer's presentation
[ ] Other (Please describe)

31. Does your Competence Committee allow for input from academic coaches?

[ ] Yes (Please describe)
[ ] No, academic coaches do not provide input
[ ] Do not have academic coaches

32. If a resident is not meeting their requirements, are they:

[ ] Required to attend the Competence Committee meeting in person
[ ] Allowed to attend the Competence Committee meeting in person
[ ] Not allowed to attend the Competence Committee meeting in Person

Decisions

33. How are decisions made to promote each resident?

[ ] Formal vote
[ ] Consensus
[ ] Other (Please describe)

34. Describe the process for communicating / reporting Competence Committee decisions to the resident?

35. Does the Competence Committee provide input into the Enhanced Educational Plan?

[ ] Yes (Please describe)
[ ] No (Please elaborate if you wish)

36. Does the Competence Committee provide input into the Remediation Plan?

[ ] Yes (Please describe)
[ ] No (Please elaborate if you wish)

37. Does the Competence Committee oversee / monitor resident remediation?

[ ] Yes (Please describe)
[ ] No (Please elaborate if you wish)

38. Is the Competence Committee involved in resident appeals?

[ ] Yes (Please describe)
[ ] No (Please elaborate if you wish)

Appendix 4 – Phase 3 Interview Guide (Two Year Mark)

Junior Members

1. Can you confirm your role on the committee (e.g., resident, newer staff, etc.) and how long you've been a member? How many meetings have you been a part of?

2. What has the experience of being a committee member been like for you so far? Probes: Are there parts of your experience that you feel have been particularly positive? Is there anything that you feel the committee could do to improve its functioning?

3. As a committee member, do you speak up when you have something to say? Do you feel like your contributions are heard? Are they influenced by other members of the committee (e.g., the Chair)? If so, how?

4. I wanted to get your thoughts on one last issue. One of the things that interests me is how CCs handle anecdotal information—things you might know or hear about a resident or their assessors that aren't necessarily documented on their assessment forms. Do you recall any situations in which this type of information has been a part of the committee's processes?

5. Do you think anecdotal information *should* be a part of the committee's processes? How should committees handle this type of information?

Senior Members

1. Can you confirm your role on the committee (e.g., chair, program director, senior staff, etc.) and how long you've been a member? How many meetings have you been a part of?

2. What has the experience of being a committee member been like for you so far? Probes: Are there parts of your experience that you feel have been particularly positive? Is there anything that you feel the committee could do to improve its functioning?

3. As a [insert role], what do you see as your role on the committee? Is there a particular perspective you bring that you feel is unique? [For Chairs only] Do you feel like each of your committee members speaks up when they have something to say or are some quieter than others? Do you employ any strategies to encourage members to voice their opinions?

4. I wanted to get your thoughts on one last issue. One of the things that interests me is how CCs handle anecdotal information—things you might know or hear about a resident or their assessors that aren't necessarily documented on their assessment forms. Do you recall any situations in which this type of information has been a part of the committee's processes?

5. Do you think anecdotal information *should* be a part of the committee's processes? How should committees handle this type of information?